(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 953 736 A1

(12) **EUROPEAN PATENT APPLICATION**
published in accordance with Art. 158(3) EPC

(43) Date of publication:
06.08.2008 Bulletin 2008/32

(21) Application number: **06812182.1**

(22) Date of filing: **30.10.2006**

(51) Int Cl.:
G10L 19/00 *(2006.01)*

(86) International application number:
PCT/JP2006/321673

(87) International publication number:
WO 2007/052612 (10.05.2007 Gazette 2007/19)

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IS IT LI LT LU LV MC NL PL PT RO SE SI SK TR

(30) Priority: **31.10.2005 JP 2005316754**
15.06.2006 JP 2006166458
02.10.2006 JP 2006271040

(71) Applicant: **MATSUSHITA ELECTRIC INDUSTRIAL CO., LTD.**
Kadoma-shi, Osaka 571-8501 (JP)

(72) Inventors:
- **GOTO, Michiyo,**
c/o Matsushita El. Ind. Co., Ltd., IPROC
Chuo-ku, Osaka-shio, Osaka 540-6207 (JP)
- **YOSHIDA, Koji,**
c/o Matsushita El. Ind. Co., Ltd., IPROC
Chuo-ku, Osaka-shi, Osaka 540-6207 (JP)
- **EHARA, Hiroyuki,**
c/o Matsushita El. Ind. Co., Ltd., IPROC
Chuo-ku, Osaka-shi, Osaka 540-6207 (JP)

(74) Representative: **Grünecker, Kinkeldey, Stockmair & Schwanhäusser**
Anwaltssozietät
Leopoldstrasse 4
80802 München (DE)

(54) STEREO ENCODING DEVICE, AND STEREO SIGNAL PREDICTING METHOD

(57) A prediction performance between the individual channels of a stereo signal is improved to improve the sound quality of a decoded signal. An LPF (101-1) interrupts the high-range component of an S1, and outputs an S1' (a low-range component). An LPF (101-2) interrupts the high-range component of an S2, and outputs an S2' (a low-range component). A prediction unit (102) predicts the S2' from the S1', and outputs a prediction parameter composed of a delay time difference (t) and an amplitude ratio (g). A first channel encoding unit (103) encodes the S1. A prediction parameter encoding unit (104) encodes the prediction parameter. The encoded parameters of the encoded parameter of the S1 and the prediction parameter are finally outputted.

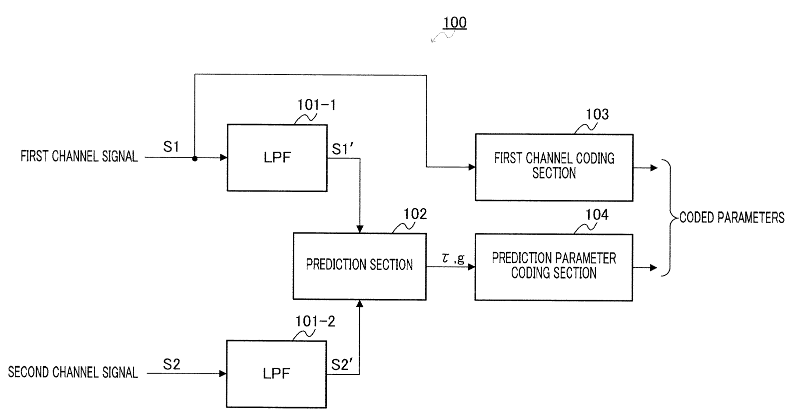

FIG.1

EP 1 953 736 A1

Description

Technical Field

[0001] The present invention relates to a stereo coding apparatus and a stereo signal prediction method.

Background Art

[0002] Monaural communication at a constant bit rate is currently mainstream in speech communication such as calls using mobile telephones in a mobile communication system. However, if transmission is realized at much higher bit rates as with the fourth-generation mobile communication system in the future, it is expected that speech communication using stereo signals having higher fidelity will be widely available.

[0003] One of coding methods for stereo speech signals is disclosed in Non-Patent Document 1. This coding method predicts one channel signal y from the other channel signal x using following equation 1 and encodes such prediction parameter a_k and d that minimize prediction errors. Here, a_k is a Kth-order prediction coefficient and d is a time difference between the two channel signals.

[1]

$$y(n)=\sum_{k=0}^{K}a_k\cdot x(n-d-k) \quad \ldots \quad \text{(Equation 1)}$$

Non-Patent Document 1: Hendrik Fuchs, "Improving Joint Stereo Audio Coding by Adaptive Inter-Channel Prediction," Applications of Signal Processing to Audio and Acoustics, Final Program and Paper Summaries, 1993 IEEE Workshop on 17-20 Oct. 1993, Page(s) 39-42.

Disclosure of Invention

Problems to be Solved by the Invention

[0004] However, in order to reduce prediction errors, with the above-described coding method, it is necessary to keep the order of a prediction coefficient at a certain order or higher, and, consequently, there is a problem that the coding bit rate increases. For example, if the order of a prediction coefficient is set a low level to lower the coding bit rate, prediction performance deteriorates and sound quality degrades auditorily.

[0005] It is therefore an object of the present invention to provide a stereo coding apparatus and stereo signal prediction method that improve prediction performance between channels of a stereo signal and improve sound quality of decoded signals.

Means for Solving the Problem

[0006] The stereo coding apparatus of the present invention employs a configuration having: a first low pass filter that lets a low-band component of a first channel signal pass; a second low pass filter that lets a low-band component of a second channel signal pass; a prediction section that predicts the low-band component of the second channel signal from the low-band component of the first channel signal and generates a prediction parameter; a first coding section that encodes the first channel signal; and a second coding section that encodes the prediction parameter.

[0007] Furthermore, the stereo signal prediction method of the present invention includes: a step of letting a low-band component of a first channel signal pass; a step of letting a low-band component of a second channel signal pass; and a step of predicting the low-band component of the second channel signal from the low-band component of the first channel signal.

Advantageous Effect of the Invention

[0008] According to the present invention, it is possible to improve prediction performance of a stereo signal between channels and improve sound quality of decoded signals.

Brief Description of Drawings

[0009]

FIG.1 is a block diagram showing the main configuration of a stereo coding apparatus according to Embodiment 1;
FIG.2A shows an example of a first channel signal;
FIG.2B shows an example of a second channel signal;
FIG.3 illustrates features of a speech signal or audio signal;
FIG.4 is a block diagram showing the main configuration of a stereo coding apparatus according to another variation of Embodiment 1;
FIG.5 is a block diagram showing the main configuration of a stereo coding apparatus according to another variation of Embodiment 1;
FIG.6 is a block diagram showing the main configuration of a stereo coding apparatus according to Embodiment 2;
FIG.7 is a block diagram showing the main configuration of a stereo coding apparatus according to Embodiment 3;
FIG.8 is a block diagram showing the main configuration of a stereo coding apparatus according to another variation of Embodiment 3;
FIG.9 is a block diagram showing the main configuration of a stereo coding apparatus according to Embodiment 4;
FIG.10 is a block diagram showing the main configuration of a stereo coding apparatus according to Embodiment 5;
FIG.11 shows an example of a cross-correlation function;
FIG.12 shows an example of a cross-correlation function;
FIG.13 is a block diagram showing the main configuration of a stereo coding apparatus according to Embodiment 6;
FIG.14 shows an example of the cross-correlation function in the case of voiced sound;
FIG.15 shows an example of the cross-correlation function in the case of unvoiced sound;
FIG.16 is a block diagram showing the main configuration of a stereo coding apparatus according to Embodiment 7;
FIG.17 shows an example of the cross-correlation function in the case of voiced sound;
FIG.18 shows an example of the cross-correlation function in the case of unvoiced sound;
FIG.19 is a block diagram showing the main configuration of a stereo coding apparatus according to Embodiment 8;
FIG.20 is a block diagram showing the main configuration of a stereo coding apparatus according to Embodiment 9;
FIG.21 shows an example of a case where a local peak of a cross-correlation function is weighted and thereby becomes a maximum cross-correlation value;
FIG.22 shows an example of a case where a maximum cross-correlation value which has not exceeded threshold φ_{th} is weighted and thereby becomes a maximum cross-correlation value exceeding threshold φ_{th}; and
FIG.23 shows an example of a case where a maximum cross-correlation value which has not exceeded threshold φ_{th} does not exceed threshold φ_{th} even after being weighted.

Best Mode for Carrying Out the Invention

[0010] Hereinafter, embodiments of the present invention will be explained below in detail with reference to the accompanying drawings.

(Embodiment 1)

[0011] FIG.1 is a block diagram showing the main configuration of stereo coding apparatus 100 according to Embodiment 1 of the present invention.

[0012] Stereo coding apparatus 100 is provided with LPF 101-1, LPF 101-2, prediction section 102, first channel coding section 103 and prediction parameter coding section 104, and receives a stereo signal comprised of a first channel signal and a second channel signal as input, performs encoding on the stereo signal and outputs coded parameters. In the present specification, a plurality of components having similar functions will be assigned the same reference numerals and further assigned different sub-numbers to distinguish from each other.

[0013] The respective sections of stereo coding apparatus 100 operate as follows.

[0014] LPF 101-1 is a low pass filter that lets only a low-band component of the input signal (original signal) pass, and more specifically, cuts off a frequency component higher than a cut-off frequency of inputted first channel signal S1, and outputs first channel signal S1' with only the low-band component remained, to prediction section 102. Likewise, LPF 101-2 also cuts off a high-band component of inputted second channel signal S2 using the same cut-off frequency as that of LPF 101-1, and outputs second channel signal S2' with only the low-band component, to prediction section 102.

[0015] Prediction section 102 predicts the second channel signal from the first channel signal using first channel signal S1' (low-band component) outputted from LPF 101-1 and second channel signal S2' (low-band component) outputted from LPF 101-2, and outputs information of this prediction (prediction parameter) to prediction parameter coding section

104. More specifically, prediction section 102 compares signal S1' with signal S2', calculates delay time difference τ between these two signals and amplitude ratio g (both are values based on the first channel signal), and outputs these values to prediction parameter coding section 104 as prediction parameters.

[0016] First channel coding section 103 carries out predetermined encoding processing on original signal S1 and outputs coded parameters obtained for the first channel. If the original signal is a speech signal, first channel coding section 103 performs encoding using, for example, a CELP (Code-Excited Linear Prediction) scheme and outputs CELP parameters such as an adaptive codebook lag and LPC coefficients as the coded parameters. On the other hand, if the original signal is an audio signal, first channel coding section 103 performs encoding using, for example, anAAC (AdvancedAudio Coding) scheme defined by MPEG-4 (Moving Picture Experts Group phase-4), and outputs the obtained coded parameters.

[0017] Prediction parameter coding section 104 applies predetermined encoding processing to the prediction parameters outputted from prediction section 102 and outputs the obtained coded parameter. For example, in the predetermined encoding processing, prediction parameter coding section 104 adopts a method of providing a codebook storing prediction parameter candidates in advance, selecting an optimum prediction parameter from this codebook and outputting an index corresponding to this prediction parameter.

[0018] Next, the above-described prediction processing carried out by prediction section 102 will be explained in further detail.

[0019] Upon calculating delay time difference τ and amplitude ratio g, prediction section 102 calculates delay time difference τ first. Delay time difference τ between low-band component S1' of the first channel signal having passed through LPF 101-1 and low-band component S2' of the second channel signal having passed through LPF 101-2 is calculated as $m=m_{max}$ that maximizes a cross-correlation function value expressed by following equation 2.

[2]

$$\phi(m) = \sum_{n=0}^{FL-1} S1'(n) \cdot S2'(n-m) \quad \ldots \text{(Equation 2)}$$

Here, n and m are sample numbers and FL is a frame length (number of samples). The cross-correlation function is obtained by shifting one signal by m and calculating a correlation value between these two signals.

[0020] Next, prediction section 102 calculates amplitude ratio g between S1' and S2' using calculated delay time difference τ obtained according to following equation 3.

[3]

$$g = \sqrt{\frac{\sum_{n=0}^{FL-1} S2'(n-\tau)^2}{\sum_{n=0}^{Fl-1} S1'(n)^2}} \quad \ldots \text{(Equation 3)}$$

Equation 3 calculates the amplitude ratio between S2' and S1' which is shifted by delay time difference τ.

[0021] Prediction section 102 predicts low-band component S2" of the second channel signal from low-band component S1' of the first channel signal using τ and g according to following equation 4.

[4]

$$S2''(n) = g \cdot S1'(n-\tau) \quad \ldots \text{(Equation 4)}$$

[0022] In this way, prediction section 102 improves the prediction performance of the stereo signal by predicting the low-band component of the second channel signal using the low-band component of the first channel signal. This principle will be explained in detail below.

[0023] FIG.2A and FIG.2B show an example of the first channel signal and the second channel signal which are the original signals. Here, for ease of explanation, an example will be explained where the number of sound sources is one.

[0024] In the first place, the stereo signal is a signal obtained by collecting sound generated from a certain source, which is common to all channels, using a plurality of (two in the present embodiment) microphones apart from each other. Therefore, when the distance from the source to the microphone becomes far, attenuation of energy of the signal becomes greater, and a delay of arrival time is brought about. Therefore, as shown in FIG.2A and FIG.2B, although the respective channels show different waveforms, signals of both channels are made more similar by correcting delay time difference Δt and amplitude difference ΔA. Here, the parameters of delay time difference and amplitude difference are characteristic parameters determined by setting positions of the microphones, and are parameters where one set of values is associated with a signal collected by one microphone.

[0025] On the other hand, as shown in FIG.3, in a speech signal or audio signal, signal energy is weighted more in the low band than the high band. Therefore, when prediction is performed as part of encoding processing, it is desirable to perform prediction by placing more importance on the low-band component than the high-band component from the standpoint of improving prediction performance.

[0026] Therefore, the present embodiment cuts off the high-band component of an input signal and calculates a prediction parameter using the remaining low-band component. The calculated coded parameter of the prediction parameter is outputted to the decoding side. That is, although the prediction parameter is calculated based on the low-band component of the input signal, this is outputted as a prediction parameter for the entire band including the high band. As described above, one set of values of a prediction parameter is associated with a signal collected by one microphone, and so, although the prediction parameter is calculated based on only the low-band component, the prediction parameter is recognized to be effective for the entire band.

[0027] Furthermore, when prediction is performed on components including even the high-band component with low energy, the prediction performance may deteriorate due to the influence of this high-band component with low accuracy. However, the present embodiment does not use the high-band component in prediction, so that the prediction performance is unlikely to deteriorate under the influence of the high-band component.

[0028] A stereo decoding apparatus according to the present embodiment that supports stereo coding apparatus 100, receives the coded parameters of the first channel outputted from first channel coding section 103, decodes these coded parameters, and thereby obtains a decoded signal of the first channel and also obtains a decoded signal of the second channel of the entire band using the codedparameter (predictionparameter) outputted from prediction parameter coding section 104 and the decoded signal of the first channel.

[0029] In this way, according to the present embodiment, a prediction parameter is calculated by cutting off the high-band component of the first channel signal in LPF 101-1, cutting off the high-band component of the second channel signal in LPF 101-2, and predicting the low-band component of the second channel signal from the low-band component of the first channel signal in prediction section 102. By outputting the coded parameter of this prediction parameter and the coded parameters of the first channel signal, it is possible to improve prediction performance of a stereo signal between the channels and improve sound quality of decoded signals. Furthermore, the high-band component of the original signal is cut off, so that it is also possible to suppress the order of the prediction coefficient to a low level.

[0030] Although a case has been described as an example with the present embodiment where first channel coding section 103 performs encoding on the first channel signal, which is an original signal, and prediction section 102 predicts second channel signal S2' from first channel signal S1', it is also possible to employ a configuration where a second channel coding section is replaced by first channel coding section 103 and encoding is applied to the second channel signal which is the original signal. In this case, prediction section 102 predicts first channel signal S1' from second channel signal S2'.

[0031] Furthermore, with the present embodiment, it is also possible to apply the above-described encoding to other input signals instead of using the first channel signal and second channel signal as input signals. FIG.4 is a block diagram showing the main configuration of stereo coding apparatus 100a according to another variation of the present embodiment. Here, first channel signal S1 and second channel signal S2 are inputted to stereo/monaural conversion section 110, and stereo/monaural conversion section 110 converts stereo signals S1 and S2 to monaural signal S_{MONO} and outputs the monaural signal.

[0032] As the conversion method in stereo/monaural conversion section 110, for example, an average signal or weighted average signal of first channel signal S1 and second channel signal S2 is obtained, and this average signal is used as monaural signal S_{MONO}. That is, the substantial coding targets in this variation are monaural signal S_{MONO} and first channel signal S1.

[0033] Therefore, LPF 111 cuts off the high-band part of monaural signal S_{MONO} and generates monaural signal S'_{MONO}, and prediction section 102a predicts first channel signal S1 from monaural signal S'_{MONO} and calculates a prediction parameter. On the other hand, monaural coding section 112 is provided instead of first channel coding section 103, and this monaural coding section 112 applies predetermined encoding processing to the monaural signal S_{MONO}.

Other operations are similar to operations of stereo coding apparatus 100.

[0034] Furthermore, the present embodiment may also be configured so as to apply smoothing processing to the prediction parameter outputted from prediction section 102. FIG.5 is a block diagram showing the main configuration of stereo coding apparatus 100b according to another variation of the present embodiment. Here, smoothing section 120 is provided after prediction section 102 which applies smoothing processing to the prediction parameter outputted from prediction section 102. Furthermore, memory 121 is provided to store the smoothed prediction parameter outputted from smoothing section 120. More specifically, smoothing section 120 applies smothing processing shown in following equations 5 and 6 using both τ(i) and g(i) of the current frame inputted from prediction section 102 and τ (i-1) and g (i-1) of the past frame inputted from memory 121, and outputs the smoothed prediction parameter to prediction parameter coding section 104b.

[5]

$$\tilde{\tau}(i) = \alpha \cdot \tilde{\tau}(i-1) + (1-\alpha) \cdot \tau(i) \quad \ldots \quad \text{(Equation 5)}$$

$$\tilde{g}(i) = \beta \cdot \tilde{g}(i-1) + (1-\beta) \cdot g(i) \quad \ldots \quad \text{(Equation 6)}$$

Here, i is a frame number, $\tilde{\tau}(i)$ and $\tilde{g}(i)$ are smoothed $\tau(i)$ and $g(i)$, and α and β are constants ranging from 0 to 1. Prediction parameter coding section 104b performs prediction on this smoothed prediction parameter using following equation 7 and calculates a prediction parameter.

[6]

$$S2''(n) = \tilde{g} \cdot S1'(n - \tilde{\tau}) \quad \ldots \quad \text{(Equation 7)}$$

Other operations are similar to operations of stereo coding apparatus 100. In this way, by smoothing variations in the values of τ and g between frames, it is possible to improve the continuity between frames of prediction signal S2" of the second channel signal.

[0035] Furthermore, although a case has been described as an example with the present embodiment where delay time difference τ and amplitude ratio g are used as prediction parameters, it is also possible to employ a configuration where the second channel signal is predicted from the first channel signal through following equation 8 using delay time difference τ and prediction coefficient series a_k instead of these parameters.

[7]

$$S2''(n) = \sum_{k=0}^{K} a_k \cdot S1'(n - \tau - k) \quad \ldots \quad \text{(Equation 8)}$$

With this configuration, it is possible to increase prediction performance.

[0036] Furthermore, although a case has been described as an example with the present embodiment where an amplitude ratio is used as one of the prediction parameters, amplitude difference, energy ratio and energy difference may also be used as parameters showing similar characteristics.

(Embodiment 2)

[0037] FIG.6 is a block diagram showing the main configuration of stereo coding apparatus 200 according to Embodiment 2 of the present invention. Stereo coding apparatus 200 has the basic configuration similar to stereo coding apparatus 100 shown in Embodiment 1, and the same components will be assigned the same reference numerals and explanations thereof will be omitted.

[0038] Stereo coding apparatus 200 is further provided with memory 201, and prediction section 202 performs different operations from prediction section 102 according to Embodiment 1 with reference to data stored in this memory 201 as appropriate.

[0039] More specifically, memory 201 accumulates prediction parameters outputted from prediction section 202 (delay time difference τ, amplitude ratio g) for predetermined past frames (N frames) and outputs the prediction parameters to prediction section 202 as appropriate.

[0040] The prediction parameters of the past frames are inputted to prediction section 202 from memory 201. Prediction section 202 determines a search range for searching a prediction parameter in the current frame according to the values of the prediction parameters of the past frames inputted from memory 201. Prediction section 202 searches a prediction parameter within the determined search range and outputs the finally obtained prediction parameter to prediction parameter coding section 104.

[0041] Explaining the above-described processing using an equation, delay time difference τ(i) of the current frame is searched within the range shown in following equation 9 assuming that the past delay time differences are τ(i-1), τ(i-2), τ(i-3), ..., τ(i-j) ..., τ(i-N).

[8]

$$\min\{\tau(i-j)\} \le \tau(i) \le \max\{\tau(i-j)\} \quad \ldots \quad \text{(Equation 9)}$$

Here, j is a value ranging from 1 to N.

[0042] Furthermore, amplitude ratio g(i) of the current frame is searched within the range shown in following equation 10 assuming that the past amplitude ratios are g(i-1), g(i-1), g(i-2), g(i-3), ..., g(i-j), ..., g(i-N).

[9]

$$\min\{g(i-j)\} \le g(i) \le \max\{g(i-j)\} \quad \ldots \quad \text{(Equation 10)}$$

Here, j is a value ranging from 1 to N.

[0043] In this way, according to the present embodiment, by determining a search range for calculating a prediction parameter based on the values of prediction parameters in the past frames, more specifically, by limiting the prediction parameter of the current frame to a value in the vicinity of the prediction parameters of the past frames, it is possible to prevent extreme prediction errors from occurring and avoid deterioration of sound quality of decoded signals.

(Embodiment 3)

[0044] FIG.7 is a block diagram showing the main configuration of stereo coding apparatus 300 according to Embodiment 3 of the present invention. Stereo coding apparatus 300 also has the basic configuration similar to stereo coding apparatus 100 shown in Embodiment 1, and the same components will be assigned the same reference numerals and explanations thereof will be omitted.

[0045] Stereo coding apparatus 300 is further provided with power detection section 301 and cut-off frequency determining section 302, and cut-off frequency determining section 302 adaptively controls cut-off frequency of LPFs 101-1 and 101-2 based on the detection result in power detection section 301.

[0046] More specifically, power detection section 301 monitors power of both first channel signal S1 and second channel signal S2 and outputs the monitoring result to cut-off frequency determining section 302. Here, a mean value for each subband is used as power.

[0047] Cut-off frequency determining section 302 averages power of first channel signal S1 for each subband over the whole band and calculates average power of the whole band. Next, cut-off frequency determining section 302 uses the calculated average power of the whole band as a threshold and compares the power of first channel signal S1 for each subband with the threshold. Cut-off frequency determining section 302 then determines cut-off frequency f1 that includes all subbands having power larger than the threshold.

[0048] Second channel signal S2 is also subjected to processing similar to that for the first channel signal S1, and cut-off frequency determining section 302 determines the value of cut-off frequency f2 of LPF 101-2. Cut-off frequency determining section 302 then determines final cut-off frequency fc common to LPFs 101-1 and 101-2 based on cut-off frequencies f1 and f2 and designates cut-off frequency fc to LPFs 101-1 and 101-2. By this means, LPFs 101-1 and 101-2 can retain all components of frequency bands having relatively large power and output such components to prediction section 102.

[0049] Normally, f1 an f2 are assumed to have the same value, and therefore cut-off frequency determining section 302 sets f1 (or f2) as final cut-off frequency fc. If f1 and f2 show different values, the cut-off frequency that allows more low-band components to remain, that is, the cut-off frequency having the greater value is adopted as fc from the standpoint of saving information safely.

[0050] In this way, according to the present embodiment, the delay time difference and amplitude ratio which are prediction parameters are calculated for signals having relatively high power, so that it is possible to improve the accuracy of calculating prediction parameters, that is, improve prediction performance.

[0051] Although an example has been described with the present embodiment where the cut-off frequency of a low pass filter is determined based on the power of the input signal, for example, the S/N ratio for each subband of an input signal may also be used. FIG.8 is a block diagram showing the main configuration of stereo coding apparatus 300a according to another variation of the present embodiment. Stereo coding apparatus 300a is provided with S/N ratio detection section 301a instead of power detection section 301 and monitors the S/N ratio for each subband of an input signal. The noise level is estimated from the input signal. Cut-off frequency determining section 302a determines a cut-off frequency of a low pass filter so as to include all subbands having relatively high S/N ratios, based on the monitoring result of S/N ratio detection section 301a. By this means, it is possible to adaptively control the cut-off frequency in a state where ambient noise exists. Thus, it is possible to calculate the delay time difference and amplitude ratio based on subbands having relatively low ambient noise level and improve the accuracy of calculating prediction parameters.

[0052] Furthermore, if the cut-off frequency per frame fluctuates discontinuously, the characteristic of a signal having passed through the low pass filter changes, and the values of τ and g also become discontinuous per frame and prediction performance deteriorates. Therefore, the cut-off frequency itself may be smoothed so that the cut-off frequency maintains continuity between frames.

(Embodiment 4)

[0053] FIG.9 is a block diagram showing the main configuration of stereo coding apparatus 400 according to Embodiment 4 of the present invention. Here, an example will be explained where an input signal is a speech signal and stereo coding apparatus 400 is a scalable coding apparatus that generates a coded parameter of a monaural signal and a coded parameter of a stereo signal.

[0054] Part of the configuration of stereo coding apparatus 400 is the same as stereo coding apparatus 100a shown in the variation of Embodiment 1 (see FIG.4, the same components will be assigned the same reference numerals). However, the input signal is speech, and, consequently, first channel coding section 410 employing a configuration different from that of stereo coding apparatus 100a is designed so that a technique of CELP coding appropriate for speech coding is applicable to first channel signal coding.

[0055] More specifically, stereo coding apparatus 400 receives a first channel signal and second channel signal as input signals, performs encoding on the monaural signal in a core layer and performs encoding on the first channel signal out of the stereo signal in an enhancement layer, and outputs both the coded parameters of the monaural signal and the coded parameters of the first channel signal to the decoding side. The decoding side can decode the second channel signal using the coded parameters of the monaural signal and the coded parameters of the first channel signal.

[0056] The core layer is provided with stereo/monaural conversion section 110, LPF 111 and monaural coding section 112, and, although this configuration is basically the same as the configuration shown with stereo coding apparatus 100a, additionally, monaural coding section 112 outputs an excitation signal of the monaural signal obtained in the middle of encoding processing to the enhancement layer.

[0057] The enhancement layer is provided with LPF 101-1, prediction section 102a, prediction parameter coding section 104 and first channel coding section 410. As in the case of Embodiment 1, prediction section 102a predicts a low-band component of the first channel signal from a low-band component of the monaural signal and outputs the generated prediction parameter to prediction parameter coding section 104 and also outputs the prediction parameter to excitation prediction section 401.

[0058] First channel coding section 410 performs encoding by separating the first channel signal into excitation information and vocal tract information. For the excitation information, excitation prediction section 401 predicts an excitation signal of the first channel signal using the prediction parameter outputted from prediction section 102a and using the excitation signal of the monaural signal outputted from monaural coding section 112. In the same way as normal CELP coding, first channel coding section 410 searches an excitation using excitation codebook 402, synthesis filter 405, distortion minimizing section 408, or the like, and obtains coded parameters of the excitation information. On the other hand, as for the vocal tract information, LPC analysis/quantization section 404 performs linear predictive analysis on the first channel signal and quantization on the analysis result, obtains a coded parameter of the vocal tract information and uses the coded parameter to generate a synthesis signal at synthesis filter 405.

[0059] In this way, according to the present embodiment, stereo/monaural conversion section 110 generates a monaural signal from the first channel signal and second channel signal, LPF 111 cuts off a high-band component of the monaural signal and generates a monaural low-band component. Prediction section 102a then predicts the low-band component of the first channel signal from the low-band component of the monaural signal through the processing similar to that in Embodiment 1, obtains a prediction parameter, performs encoding on the first channel signal using the prediction parameter according to a method compatible with CELP coding and obtains coded parameters of the first channel signal. The coded parameters of this first channel signal together with the coded parameters of the monaural signal are outputted to the decoding side. With this configuration, it is possible to realize a monaural-stereo scalable coding apparatus, improve prediction performance of a stereo signal between channels and improve sound quality of decoded signals.

(Embodiment 5)

[0060] FIG.10 is a block diagram showing the main configuration of stereo coding apparatus 500 according to Embodiment 5 of the present invention. Stereo coding apparatus 500 also has the basic configuration similar to that of stereo coding apparatus 100 shown in Embodiment 1, and the same components will be assigned the same reference numerals and explanations thereof will be omitted.

[0061] Stereo coding apparatus 500 is provided with threshold setting section 501 and prediction section 502, and prediction section 502 decides the reliability of this cross-correlation function by comparing threshold φ_{th} preset in threshold setting section 501 with the value of cross-correlation function φ.

[0062] More specifically, prediction section 502 calculates cross-correlation function φ expressed by following equation 11 using low-band component S1' of the first channel signal having passed through LPF 101-1 and low-band component S2' of the second channel signal having passed through LPF 101-2,

[10]

$$\phi(m) = \frac{\sum_{n=0}^{FL-1} S1'(n) \cdot S2'(n-m)}{\sqrt{\sum_{n=0}^{FL-1} S1'(n)^2}\sqrt{\sum_{n=0}^{FL-1} S2'(n-m)^2}} \quad \dots \text{ (Equation 11)}$$

where, cross-correlation function φ is assumed to be normalized with the autocorrelation function of each channel signal. Furthermore, n and m are sample numbers and FL is a frame length (number of samples). As is apparent from equation 11, the maximum value of φ is 1.

[0063] Prediction section 502 then compares threshold φ_{th} preset in threshold setting section 501 with the maximum value of cross-correlation function φ and, when this is equal to or greater than the threshold, decides that this cross-correlation function is reliable. In other words, prediction section 502 compares threshold φ_{th} preset in threshold setting section 501 with sample values of cross-correlation function φ, and, when there is at least one sample point which is equal to or greater than the threshold, decides that this cross-correlation function is reliable. FIG.11 shows an example of cross-correlation function φ. This is an example where the maximum value of the cross-correlation function exceeds the threshold.

[0064] In such a case, prediction section 502 calculates delay time difference τ between low-band component S1' of the first channel signal and low-band component S2' of the second channel signal as $m=m_{max}$ that maximizes the value of the cross-correlation function expressed by above-described equation 11.

[0065] On the other hand, when the maximum value of cross-correlation function φ does not reach threshold φ_{th}, prediction section 502 determines delay time difference τ already determined in the previous frame as delay time difference τ of the frame. FIG. 12 also shows an example of cross-correlation function φ. Here, an example is shown where the maximum value of the cross-correlation function does not exceed the threshold.

[0066] Prediction section 502 calculates amplitude ratio g using a method similar to that of Embodiment 1.

[0067] In this way, according to the present embodiment, to calculate delay time difference τ with high reliability, whether or not the value of the cross-correlation function is reliable is decided, and then the value of delay time difference τ is determined. More specifically, the cross-correlation function normalized with the autocorrelation function of each channel signal is used as the cross-correlation function upon calculating the delay time difference, a threshold is provided in advance, and, when the maximum value of the cross-correlation function is equal to or greater than the threshold, $m=m_{max}$ that maximizes the value of the cross-correlation function is determined as the delay time difference. On the other hand, when the cross-correlation function does not reach the threshold at all, the delay time difference determined in the previous frame is determined as the delay time difference of the frame. With this configuration, it is possible to calculate a delay time difference accurately.

(Embodiment 6)

[0068] FIG.13 is a block diagram showing the main configuration of stereo coding apparatus 600 according to Embodiment 6 of the present invention. Stereo coding apparatus 600 has the basic configuration similar to that of stereo coding apparatus 500 shown in Embodiment 5, and the same components will be assigned the same reference numerals and explanations thereof will be omitted.

[0069] Stereo coding apparatus 600 is further provided with voiced/unvoiced sound decision section 601, which decides whether a first channel signal and a second channel signal not having passed through low pass filters are voiced sound or unvoiced sound to set a threshold in threshold setting section 501.

[0070] More specifically, voiced/unvoiced sound decision section 601 calculates the value of autocorrelation function φ_{ss} using first channel signal S1 and second channel signal S2 according to following equation 12. [11]

$$\phi_{SS}(m) = \frac{\sum_{n=0}^{FL-1} S(n) \cdot S(n-m)}{\sqrt{\sum_{n=0}^{FL-1} S(n)^2}\sqrt{\sum_{n=0}^{FL-1} S(n-m)^2}} \quad \ldots \text{(Equation 12)}$$

Here, S (n) is a first channel signal or second channel signal, n and m are sample numbers and FL is a frame length (number of samples). As is apparent from equation 12, the maximum value of φ_{ss} is 1.

[0071] A threshold for deciding voiced/unvoiced sound is preset in voiced/unvoiced sound decision section 601. Voiced/unvoiced sound decision section 601 compares the value of autocorrelation function φ_{ss} of the first channel signal or second channel signal with the threshold, decides that the signal is a voiced sound when the value exceeds the threshold and decides that the signal is not a voiced sound (that is, an unvoiced sound) when the value does not exceed the threshold. That is, a decision on voiced/unvoiced sound is made for both the first channel signal and second channel signal. Voiced/unvoiced sound decision section 601 then takes into consideration the values of autocorrelation function φ_{ss} of the first channel signal and autocorrelation function φ_{ss} of the second channel signal by, for example, calculating a mean value thereof and decides whether these channel signals are voiced or unvoiced sounds. The decision result is outputted to threshold setting section 501.

[0072] Threshold setting section 501 changes the threshold setting depending on whether the channel signals are decided as voiced or not decided as voiced sound. More specifically, threshold setting section 501 sets threshold φ_V used in the case of voiced sound smaller than threshold φ_{UV} used in the case of unvoiced sound. The reason is that periodicity exists in the case of voiced sound, and, consequently, there is a large difference between the value of the cross-correlation function which has a local peak and other values of the cross-correlation function which do not have local peaks. On the other hand, no periodicity exists in the case of unvoiced sound (because it is noise-like sound), and, consequently, the difference between the value of the cross-correlation function which has a local peak and other values of the cross-correlation function which do not have local peaks is not large.

[0073] FIG.14 shows an example of the cross-correlation function in the case of voiced sound. Furthermore, FIG.15 shows an example of the cross-correlation function in the case of unvoiced sound. Both figures show the threshold as well. As shown in this figure, the cross-correlation function has different aspects between voiced sound and unvoiced sound, and, consequently, a threshold is set so as to adopt a value of a reliable cross-correlation function, and the method of setting the threshold is changed depending on whether a signal has a voiced sound property or an unvoiced sound property. That is, by setting a greater threshold of the cross-correlation function for a signal judged to have an unvoiced sound property, the signal is not adopted as a delay time difference unless there is a large difference between the value of the cross-correlation function and values of other cross-correlation functions which do not become local peaks, so that it is possible to improve the reliability of the cross-correlation function.

[0074] In this way, according to the present embodiment, by deciding voiced/unvoiced sound using the first channel signal and second channel signal not having passed through the low pass filter, the threshold for deciding the reliability of the cross-correlation function is changed depending on whether the signal is a voiced sound or unvoiced sound. More specifically, a smaller threshold is set for voiced sound than for unvoiced sound. Therefore, it is possible to determine the delay time difference more accurately.

(Embodiment 7)

[0075] FIG.16 is a block diagram showing the main configuration of stereo coding apparatus 700 according to Embodiment 7 of the present invention. Stereo coding apparatus 700 has the basic configuration similar to that of stereo coding apparatus 600 shown in Embodiment 6, and the same components will be assigned the same reference numerals and explanations thereof will be omitted.

[0076] Stereo coding apparatus 700 is provided with coefficient setting section 701, threshold setting section 702, and prediction section 703 after voiced/unvoiced sound decision section 601, and multiplies a maximum value of a cross-correlation function by a coefficient according to a voiced/unvoiced decision result and determines a delay time difference using the maximum value of the cross-correlation function having multiplied by this coefficient.

[0077] More specifically, coefficient setting section 701 sets coefficient g which varies depending on whether the signal is voiced or unvoiced sound based on the decision result outputted from voiced/unvoiced sound decision section 601 and outputs coefficient g to threshold setting section 702. Here, coefficient g is set a positive value less than 1 based on the maximum value of the cross-correlation function. Furthermore, greater coefficient g_V is set in the case of voiced sound than coefficient g_{UV} in the case of unvoiced sound. Threshold setting section 702 sets a value obtained by multiplying maximum value φ_{max} of the cross-correlation function by coefficient g as threshold φ_{th} and outputs the set value to prediction section 703. Prediction section 703 detects local peaks whose apices are included in the area between this threshold φ_{th} and maximum value φ_{max} of the cross-correlation function.

[0078] FIG.17 shows an example of the cross-correlation function in the case of voiced sound. Furthermore, FIG.18 shows an example of the cross-correlation function in the case of unvoiced sound. Both figures show thresholds as well. Prediction section 703 detects local peaks of the cross-correlation function whose apices exist in the area between maximum value φ_{max} and threshold φ_{th}, and, unless local peaks other than the peaks (encircled peaks in the figure) showing maximum values are detected, decides $m=m_{max}$ that maximizes the value of the cross-correlation function as a delay time difference. For example, in the example of FIG.17, only one local peak exists in the area between φ_{max} and φ_{th}, and $m=m_{max}$ is adopted as delay time difference τ. On the other hand, if local peaks other than the peaks showing the maximum values are detected, the delay time difference of the previous frame is determined as the delay time difference of the frame. For example, in the example of FIG.18, four local peaks (encircled peaks in the figure) exist in the area between φ_{max} and φ_{th}, and, consequently, $m=m_{max}$ is not adopted as delay time difference τ and the delay time difference of the previous frame is adopted as the delay time difference of the frame.

[0079] The reason for setting different thresholds by changing the coefficient between voiced sound and unvoiced sound, is that there is periodicity in the case of voiced sound, which causes a large difference between the value of the cross-correlation function which normally has a local peak and other values of the cross-correlation function which do not have local peaks, and therefore only the vicinity of maximum value φ_{max} needs to be checked. On the other hand, in the case of unvoiced sound, there is no periodicity (noise-like sound), the difference between the value of the cross-correlation function which has a local peak and other values of the cross-correlation function which do not have local peaks is not large, and therefore it is necessary to check whether or not there is a sufficient difference between maximum value φ_{max} and other local peaks.

[0080] In this way, according to the present embodiment, a maximum value of the cross-correlation function is used as a standard and a value obtained by multiplying the maximum value by a positive coefficient less than 1 is used as a threshold. Here, the value of the coefficient to be multiplied varies depending on whether the signal is voiced or unvoiced sound (the value is made greater for voiced sound than for unvoiced sound). Local peaks existing between the maximum value of the cross-correlation function and the threshold are detected, and, if any local peak other than the peak showing the maximum value is not detected, the value of $m=m_{max}$ that maximizes the value of the cross-correlation function is determined as the delay time difference. On the other hand, if any local peak other than the peak showing the maximum value is detected, the delay time difference of the previous frame is determined as the delay time difference of the frame. That is, based on the maximum value of the cross-correlation function, the delay time difference is set according to the number of local peaks included in a predetermined range from the maximum value of the cross-correlation function. The delay time difference can be determined accurately by employing such a configuration.

(Embodiment 8)

[0081] FIG.19 is a block diagram showing the main configuration of stereo coding apparatus 800 according to Embodiment 8 of the present invention. Stereo coding apparatus 800 has the basic configuration similar to that of stereo coding apparatus 500 shown in Embodiment 5, and the same components will be assigned the same reference numerals and explanations thereof will be omitted.

[0082] Stereo coding apparatus 800 is further provided with cross-correlation function value storage section 801, and prediction section 802 performs different operations from prediction section 502 according to Embodiment 5 with reference to cross-correlation function values stored in this cross-correlation function value storage section 801.

[0083] More specifically, cross-correlation function value storage section 801 accumulates smoothed maximum cross-correlation values outputted from prediction section 802 and outputs the maximum cross-correlation values to prediction section 802 as appropriate.

[0084] Prediction section 802 compares threshold φ_{th} preset in threshold setting section 501 with the maximum value of cross-correlation function φ, and, when this is equal to or greater than the threshold, decides that this cross-correlation function is reliable. In other words, prediction section 802 compares threshold φ_{th} preset in threshold setting section 501 with sample values of cross-correlation function φ, and, when there is at least one sample point which is equal to or greater than the threshold, decides that this cross-correlation function is reliable.

[0085] In such a case, prediction section 802 calculates delay time difference τ between low-band component S1' of a first channel signal and low-band component S2' of a second channel signal as $m=m_{max}$ that maximizes the value of the cross-correlation function expressed by equation 12 described above.

[0086] On the other hand, when the maximum value of cross-correlation function φ does not reach threshold φ_{th}, prediction section 802 determines delay time difference τ using the smoothed maximum cross-correlation value of the previous frame outputted from cross-correlation function value storage section 801. The smoothed maximum cross-correlation value is expressed by following equation 13.

[12]

$$\phi_{smooth} = \phi_{smooth\_prev} \cdot \alpha + \phi_{\max} \cdot (1-\alpha) \quad \ldots \quad \text{(Equation 13)}$$

Here, φ_{smooth_prev} is a smoothed maximum cross-correlation value of the previous frame, φ_{max} is a maximum cross-correlation value of the current frame and α is a smoothing coefficient and a constant that satisfies $0<\alpha<1$.

[0087] Further, smoothed maximum cross-correlation values accumulated in cross-correlation function value storage section 801 are used as φ_{smooth_prev} upon determining the delay time difference of the next frame.

[0088] More specifically, when the maximum value of cross-correlation function φ does not reach threshold φ_{th}, prediction section 802 compares smoothed maximum cross-correlation value φ_{smooth_prev} of the previous frame with preset threshold $\varphi_{th\_smoth\_prev}$. As a result, when φ_{smooth_prev} is greater than $\varphi_{th\_smooth\_prev}$, the delay time difference of the previous frame is determined as delay time difference τ of the current frame. On the contrary, when φ_{smooth_prev} does not exceed $\varphi_{th\_smooth\_prev}$, the delay time difference of the current frame is set 0.

[0089] Prediction section 802 calculates amplitude ratio g using a method similar to that of Embodiment 1.

[0090] In this way, according to the present embodiment, when the maximum cross-correlation value of the current frame is low, the obtained delay time difference has also low reliability, and, consequently, by using as a substitute, a delay time difference of the previous frame having higher reliability decided using the smoothed maximum cross-correlation value in the previous frame, it is possible to determine the delay time difference more accurately.

(Embodiment 9)

[0091] FIG.20 is a block diagram showing the main configuration of stereo coding apparatus 900 according to Embodiment 9 of the present invention. Stereo coding apparatus 900 has the basic configuration similar to that of stereo coding apparatus 600 shown in Embodiment 6, and the same components will be assigned the same reference numerals and explanations thereof will be omitted.

[0092] Stereo coding apparatus 900 is further provided with weight setting section 901 and delay time difference storage section 902, and weight setting section 901 outputs weights according to voiced/unvoiced sound decision result of a first channel signal and second channel signal, and prediction section 903 performs different operations from prediction section 502 according to Embodiment 6 using this weight and the delay time difference stored in delay time difference storage section 902.

[0093] Weight setting section 901 changes weight w (>1.0) depending on whether voiced/unvoiced sound decision section 601 decides voiced sound or unvoiced sound. More specifically, weight setting section 901 sets larger weight w in the case of unvoiced sound than weight w in the case of voiced sound.

[0094] The reason is that, in the case of voiced sound, there is periodicity, and so the difference between the maximum value of the cross-correlation function and other values of the cross-correlation function at local peaks is relatively large and the amount of shift showing the maximum cross-correlation value shows a correct delay difference with high reliability, while, in the case of unvoiced sound, there is no periodicity (noise-like sound), and so the difference between the maximum value of the cross-correlation function and other values of the cross-correlation function at local peaks is relatively small, and the amount of shift showing the maximum cross-correlation value does not always show a correct delay difference. Therefore, a more accurate delay difference can be obtained by setting larger weight w in the case of unvoiced sound and making the delay difference of the previous frame easier to select.

[0095] Delay time difference storage section 902 accumulates delay time difference τ outputted from prediction section 903 and outputs this to prediction section 903 as appropriate.

[0096] Prediction section 903 determines a delay difference using weight w set by weight setting section 901 as follows. First, a candidate of delay time difference τ between low-band component S1' of the first channel signal having passed through LPF 101-1 and low-band component S2' of the second channel signal having passed through LPF 101-2 is determined as $m=m_{max}$ that maximizes the value of the cross-correlation function expressed by equation 11 above. The cross-correlation function is normalized with the autocorrelation function of each channel signal.

[0097] In equation 11, n is a sample number and FL is a frame length (number of samples). Furthermore, m is the amount of shift.

[0098] Here, when the difference between the value of m and the value of the delay time difference of the previous frame stored in delay time difference storage section 902 is within a preset range, prediction section 903 multiplies the cross-correlation value obtained by equation 11 described above by the weight set by weight setting section 901 as shown in following equation 14. The preset range is set based on delay time difference τ_{prev} in the previous frame stored in delay time difference storage section 902.

[13]

$$\phi_w(m) = w \times \phi(m) \quad \ldots \text{ (Equation 14)}$$

On the other hand, when the value of m is outside the preset range, the expression becomes as following equation 15.
[14]

$$\phi_w(m) = \phi(m) \quad \ldots \text{ (Equation 15)}$$

The reliability of the candidate of the delay time difference τ obtained in this way is judged by maximum value (maximum cross-correlation value) φ_{max} of the cross-correlation function expressed by above-described equation 14 and above-described equation 15 and final delay time difference τ is determined. More specifically, threshold φ_{th} preset in threshold setting section 501 is compared with maximum cross-correlation value φ_{max}, and, if maximum cross-correlation value φ_{max} is equal to or greater than threshold φ_{th}, this cross-correlation function is judged to be reliable, and $m=m_{max}$ that maximizes the value of the cross-correlation function is determined as delay time difference τ.

[0099] FIG.21 shows an example of a case where a local peak of the cross-correlation function is weighted and thereby becomes a maximum cross-correlation value. Furthermore, FIG.22 shows an example of a case where a maximum cross-correlation value which has not exceeded threshold φ_{th} is weighted and thereby becomes a maximum cross-correlation value that exceeds threshold φ_{th}. Furthermore, FIG.23 shows an example of a case where a maximum cross-correlation value which has not exceeded threshold φ_{th} is weighted and still does not exceed threshold φ_{th}. In the case shown in FIG.23, the delay time difference of the current frame is set 0.

[0100] In this way, according to the present embodiment, when the difference between amount of shift m of a sample and the delay time difference of the previous frame is within a predetermined range, by weighting the cross-correlation function value, the cross-correlation function value with the amount of shift near the delay time difference of the previous frame is evaluated as a relatively greater value than the cross-correlation function value of other amounts of shift, and the amount of shift near the delay time difference of the previous frame is selected more easily, so that it is possible to calculate the delay time difference in the current frame more accurately.

[0101] Although a configuration has been described with the present embodiment where the weight by which the cross-correlation function value is multiplied varies according to the voiced/unvoiced sound decision result, a configuration may be employed where the cross-correlation function value is always multiplied by a fixed weight regardless of the voiced/unvoiced sound decision result.

[0102] Further, although examples have been described with Embodiment 5 to Embodiment 9 where processing on the first channel signal and second channel signal having passed through low pass filters, the processing of Embodiment 5 to Embodiment 9 may also be applied to signals not subjected to low pass filter processing.

[0103] Furthermore, instead of the first channel signal and second channel signal having passed through low pass filters, a residual signal (excitation signal) of the first channel signal having passed through the low pass filter and a residual signal (excitation signal) of the second channel signal having passed through the low pass filter may also be used.

[0104] Furthermore, instead of the first channel signal and second channel signal not subjected to low pass filter processing, the residual signal (excitation signal) of the first channel signal and the residual signal (excitation signal) of the second channel signal may also be used.

[0105] Embodiments of the present invention have been explained above.

[0106] The stereo coding apparatus and stereo signal prediction method according to the present invention are not limited to the above-described embodiments, but can be implemented with various modifications. For example, above-described embodiments may be implemented in combination as appropriate.

[0107] The stereo speech coding apparatus according to the present invention can be provided to communication terminal apparatuses and base station apparatuses in a mobile communication system, so that it is possible to provide a communication terminal apparatus, base station apparatus and mobile communication system having operational effects similar to those described above.

[0108] Although a case has been described with the above embodiments as an example where the present invention is implemented with hardware, the present invention can be implemented with software. For example, by describing the stereo coding method and stereo decoding method algorithm according to the present invention in a programming language, storing this program in a memory and making the information processing section execute this program, it is possible to implement the same function as the stereo coding apparatus and stereo decoding apparatus of the present invention.

[0109] Furthermore, each function block employed in the description of each of the aforementioned embodiments may typically be implemented as an LSI constituted by an integrated circuit. These may be individual chips or partially or totally contained on a single chip.

[0110] "LSI" is adopted here but this may also be referred to as "IC," "system LSI," "super LSI," or "ultra LSI" depending on differing extents of integration.

[0111] Further, the method of circuit integration is not limited to LSI's, and implementation using dedicated circuitry or general purpose processors is also possible. After LSI manufacture, utilization of an FPGA (Field Programmable Gate Array) or a reconfigurable processor where connections and settings of circuit cells in an LSI can be reconfigured is also possible.

[0112] Further, if integrated circuit technology comes out to replace LSI's as a result of the advancement of semiconductor technology or a derivative other technology, it is naturally also possible to carry out function block integration using this technology. Application of biotechnology is also possible.

[0113] The present application is based on Japanese Patent Application No.2005-316754, filed on October 31, 2005, Japanese Patent Application No.2006-166458, filed on June 15, 2006 and Japanese Patent Application No.2006-271040, filed on October 2, 2006, the entire content of which is expressly incorporated by reference herein.

Industrial Applicability

[0114] The stereo coding apparatus and stereo signal prediction method according to the present invention are applicable to, for example, communication terminal apparatuses, base station apparatuses in a mobile communication system.

Claims

1. A stereo coding apparatus comprising:

a first lowpass filter that lets a low-band component of a first channel signal pass;
a second low pass filter that lets a low-band component of a second channel signal pass;
a prediction section that predicts the low-band component of the second channel signal from the low-band component of the first channel signal and generates a prediction parameter;
a first coding section that encodes the first channel signal; and
a second coding section that encodes the prediction parameter.

2. The stereo coding apparatus according to claim 1, wherein the prediction section performs the prediction and generates information of a delay time difference and an amplitude ratio between the low-band component of the first channel signal and the low-band component of the second channel signal.

3. The stereo coding apparatus according to claim 1, further comprising a memory that stores the prediction parameter, wherein the prediction section generates a prediction parameter within a predetermined range using the prediction parameter based on the past prediction parameter stored in the memory.

4. The stereo coding apparatus according to claim 1, further comprising:

an acquisition section that acquires power of the first channel signal and the second channel signal; and
a determination section that determines cut-off frequencies of the first low pass filter and the second low pass filter based on the power of the first channel signal and the second channel signal.

5. The stereo coding apparatus according to claim 1, further comprising:

a detection section that detects signal to noise ratios of the first channel signal and the second channel signal; and
a determination section that determines cut-off frequencies of the first low pass filter and the second low pass filter based on the signal to noise ratios of the first channel signal and the second channel signal.

6. A stereo coding apparatus comprising:

a conversion section that converts a first channel signal and a second channel signal to monaural signals;
a first lowpass filter that lets a low-band component of the monaural signal pass;
a second low pass filter that lets a low-band component of the first channel signal pass;
a prediction section that predicts the low-band component of the first channel signal from the low-band component of the monaural signal and generates a prediction parameter;
a first coding section that encodes the monaural signal; and
a second coding section that encodes the first channel signal using the prediction parameter.

7. The stereo coding apparatus according to claim 6, wherein the second coding section encodes the first channel signal separated into excitation information and vocal tract information and uses the prediction parameter for encoding the excitation information.

8. The stereo coding apparatus according to claim 1, further comprising a smoothing section that smoothes the prediction parameter,
wherein the second coding section encodes the smoothed prediction parameter.

9. The stereo coding apparatus according to claim 2, further comprising a calculation section that mutually shifts the low-band component of the first channel signal and the low-band component of the second channel signal, and calculates a value of a cross-correlation function of these two signals,
wherein, upon generating information of the delay time difference, the prediction section sets an amount of shift that maximizes the cross-correlation function as a delay time difference, when the value of the cross-correlation function is equal to or greater than a threshold, and uses the delay time difference of the previous frame again when the value of the cross-correlation function is less than the threshold.

10. The stereo coding apparatus according to claim 9, further comprising a decision section that makes a voiced/unvoiced sound decision on the first channel signal and the second channel signal,
wherein the prediction section sets the threshold based on the decision result in the decision section.

11. The stereo coding apparatus according to claim 9, wherein, if a maximum value of the cross-correlation function is equal to or greater than a first threshold, the prediction section sets an amount of shift that maximizes the cross-correlation function as the delay time difference, and, if the maximum value of the cross-correlation function is less than the first threshold and the maximum value of the smoothed cross-correlation value of the previous frame is equal to or greater than a second threshold, the prediction section sets the delay time difference of the previous frame as the delay time difference of a current frame, and, if the maximum value of the smoothed cross-correlation value of the previous frame is less than the second threshold, the prediction section sets the delay time difference of the current frame as 0.

12. The stereo coding apparatus according to claim 9, wherein, when the difference between the delay time difference of the previous frame and the amount of shift of a sample upon mutually shifting the low-band component of the first channel signal and the low-band component of the second channel signal is within a predetermined range, the prediction section assigns a weight to the value of the cross-correlation function.

13. The stereo coding apparatus according to claim 12, further comprising:

a decision section that makes a voiced/unvoiced sound decision on the first channel signal and the second channel signal; and
a weight setting section that sets the weight based on the decision result in the decision section.

14. The stereo coding apparatus according to claim 2, further comprising:

a decision section that makes a voiced/unvoiced sound decision on the first channel signal and the second channel signal; and
a calculation section that mutually shifts the low-band component of the first channel signal and the low-band component of the second channel signal and calculates a value of a cross-correlation function of these two signals,

wherein, upon generating information of the delay time difference, the prediction section sets the delay time difference according to the number of local peaks included within a predetermined range from a maximum value of the cross-correlation function.

15. A communication terminal apparatus comprising the stereo coding apparatus according to claim 1.

16. A base station apparatus comprising the stereo coding apparatus according to claim 1.

17. A stereo signal prediction method comprising:

a step of letting a low-band component of a first channel signal pass;
a step of letting a low-band component of a second channel signal pass; and
a step of predicting the low-band component of the second channel signal from the low-band component of the first channel signal.

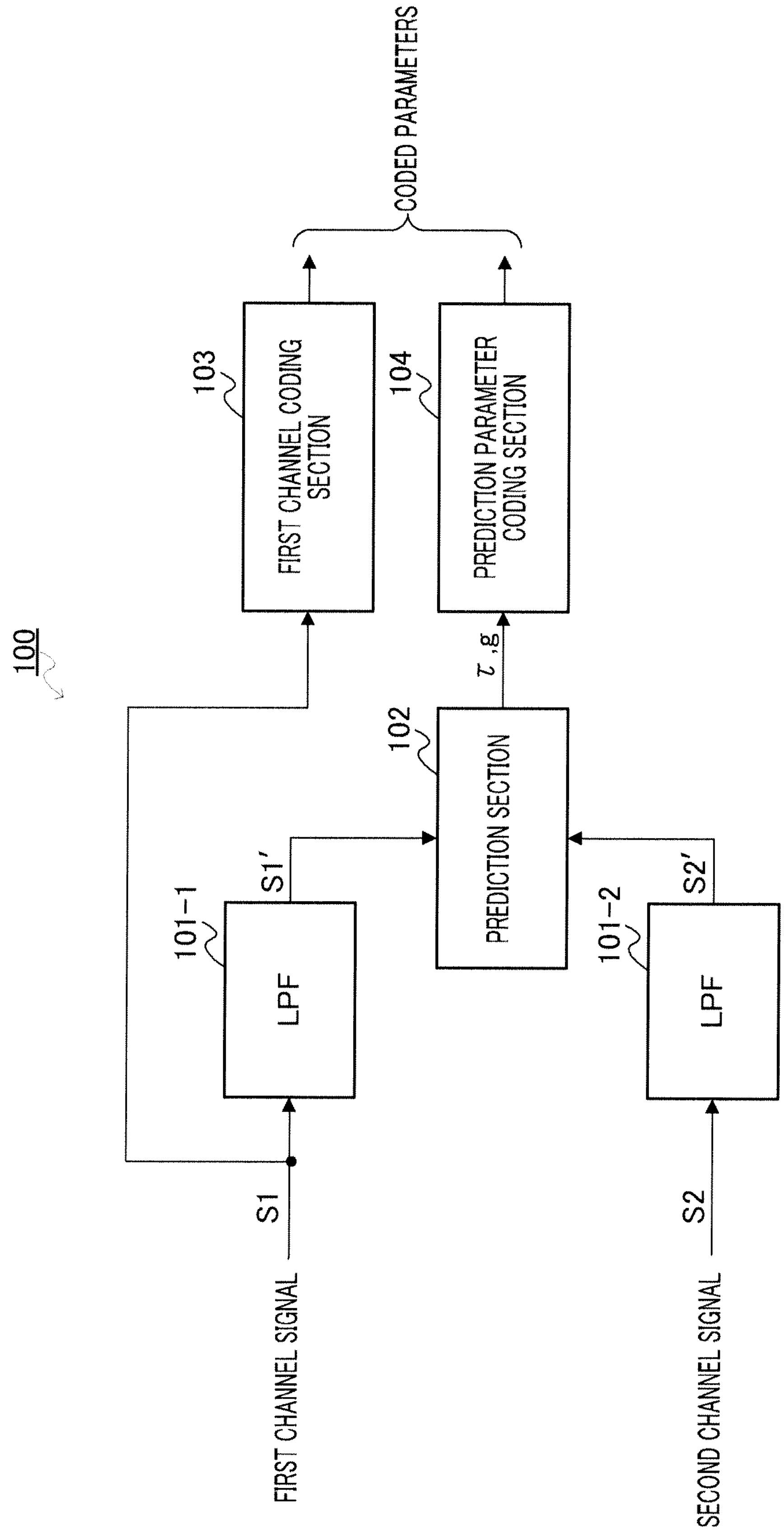
100
FIRST CHANNEL SIGNAL
S1
101-1
LPF
S1'
SECOND CHANNEL SIGNAL
S2
101-2
LPF
S2'
102
PREDICTION SECTION
τ ,g
103
FIRST CHANNEL CODING SECTION
104
PREDICTION PARAMETER CODING SECTION
CODED PARAMETERS

FIG.1

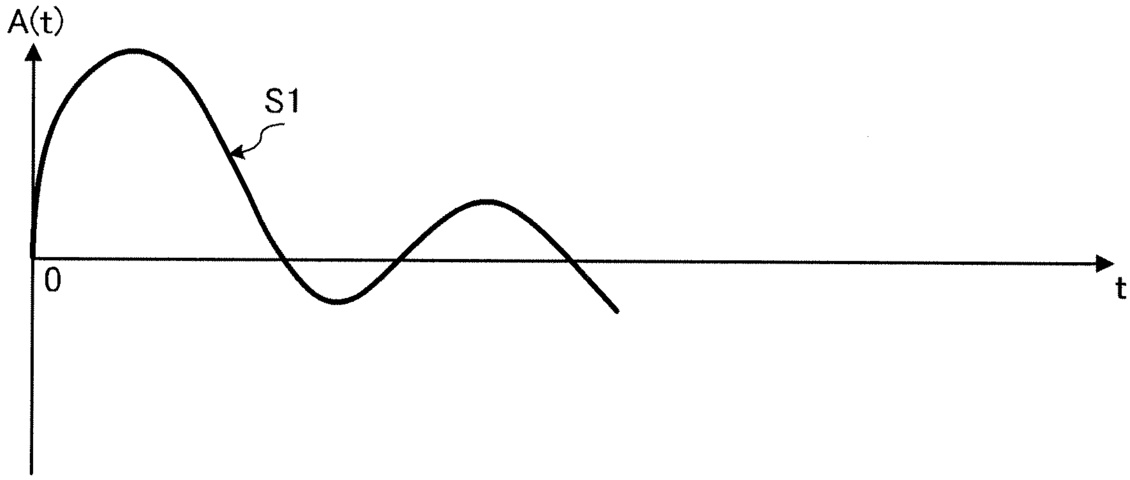
A(t)
S1
0
t

FIG.2A

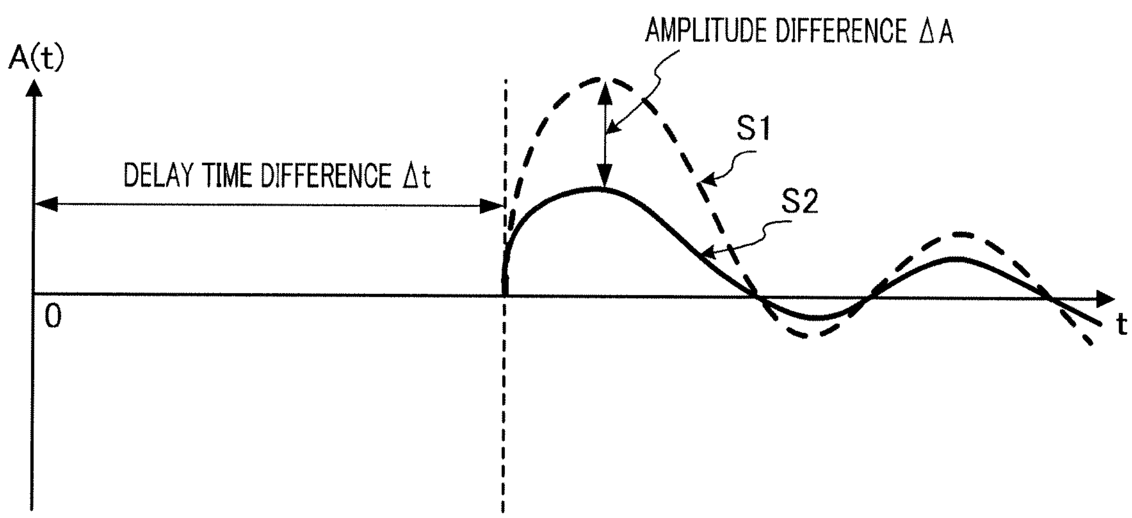
AMPLITUDE DIFFERENCE ΔA
A(t)
S1
DELAY TIME DIFFERENCE Δt
S2
0
t

FIG.2B

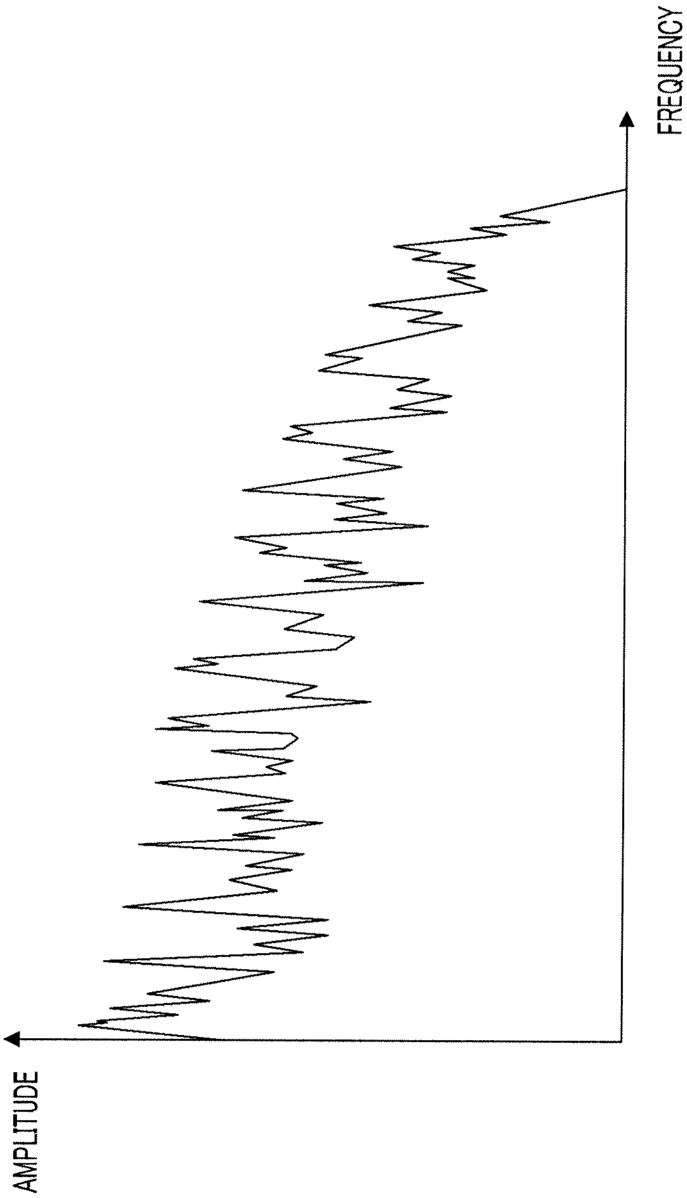

FIG.3

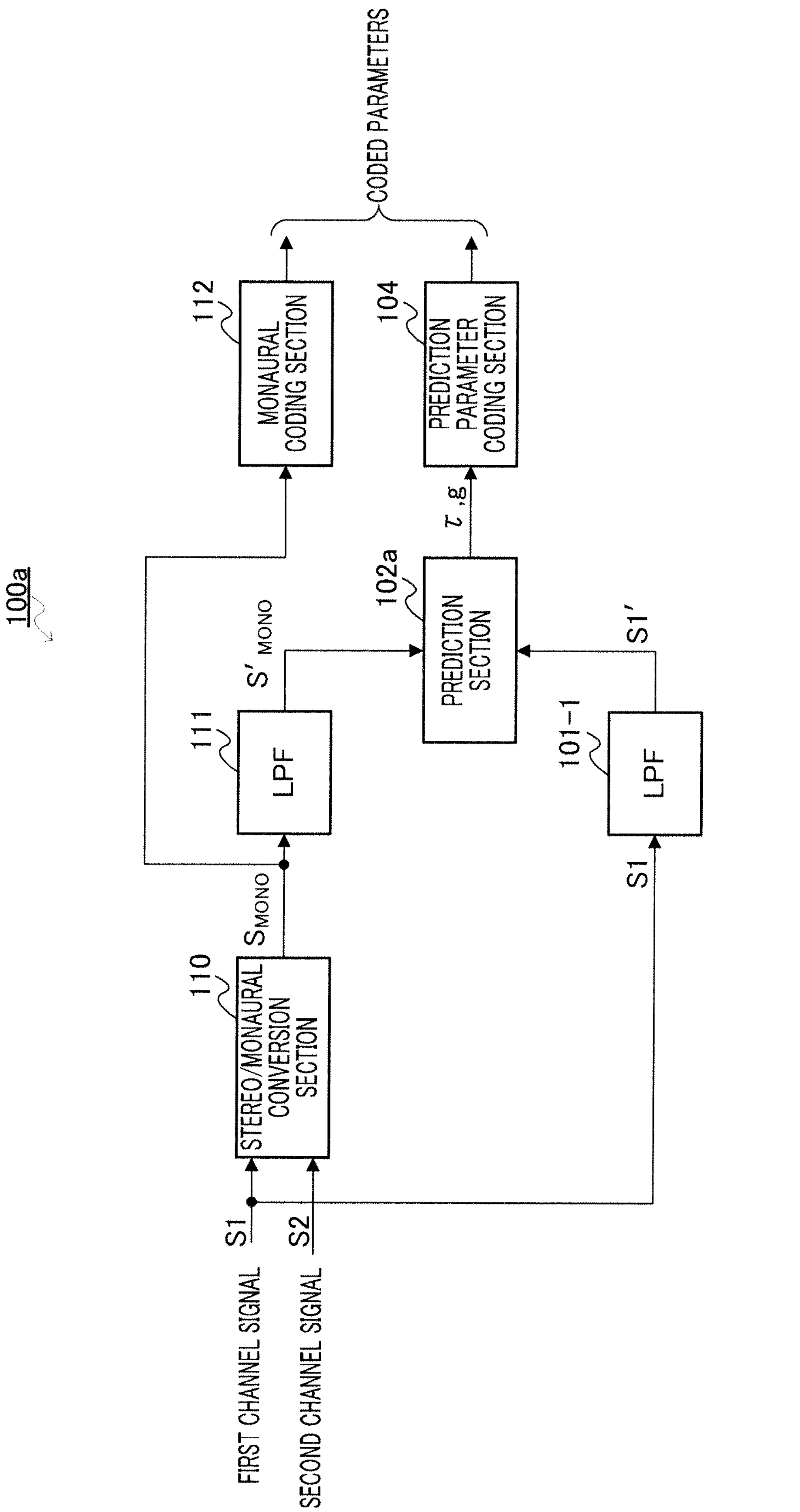
100a
FIRST CHANNEL SIGNAL
SECOND CHANNEL SIGNAL
S1
S2
110
STEREO/MONAURAL CONVERSION SECTION
S_MONO
111
LPF
S'_MONO
101-1
LPF
S1
S1'
102a
PREDICTION SECTION
τ, g
112
MONAURAL CODING SECTION
104
PREDICTION PARAMETER CODING SECTION
CODED PARAMETERS

FIG.4

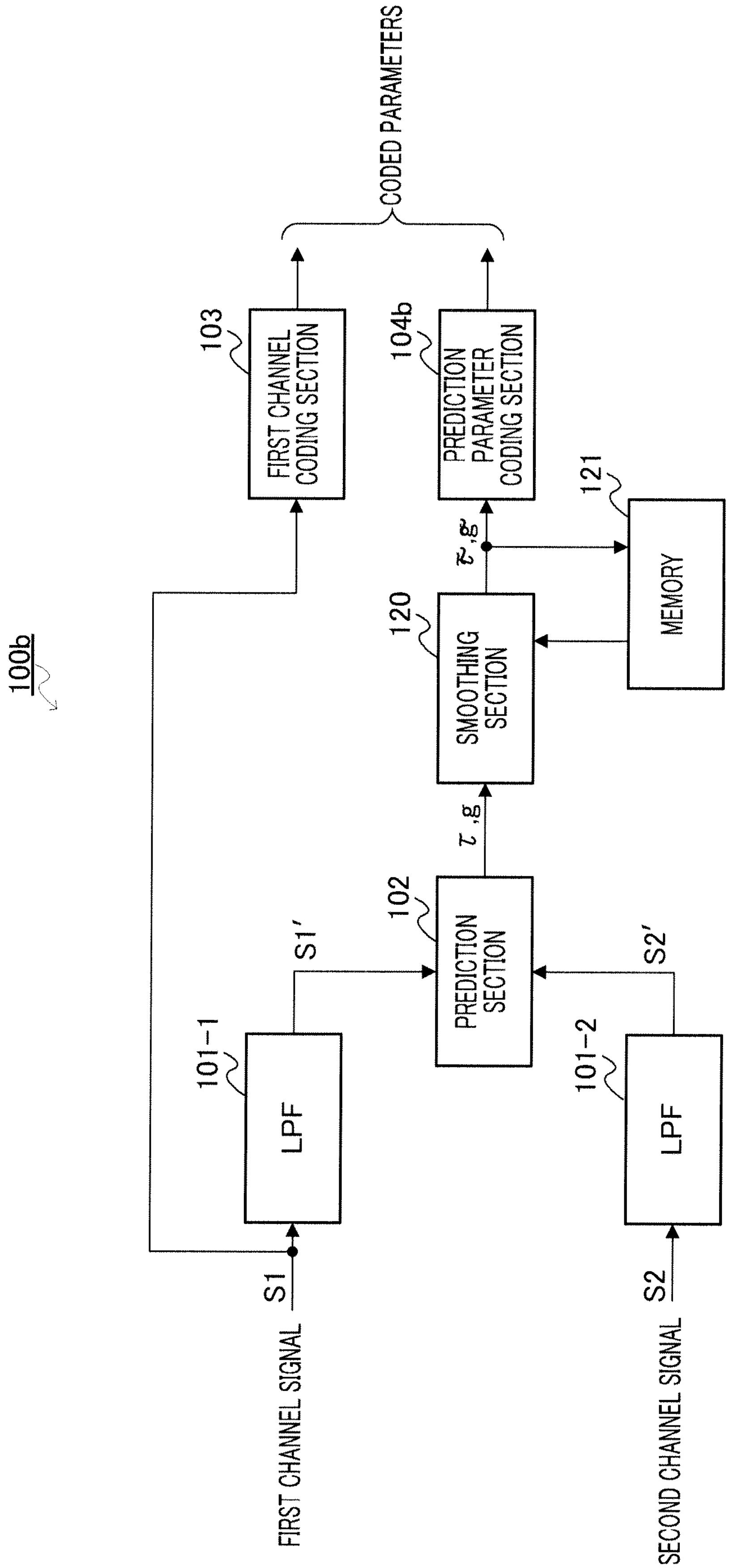
100b
FIRST CHANNEL SIGNAL
S1
SECOND CHANNEL SIGNAL
S2
101-1
LPF
101-2
LPF
S1′
S2′
102
PREDICTION SECTION
τ,g
120
SMOOTHING SECTION
121
MEMORY
τ̃,g̃
103
FIRST CHANNEL CODING SECTION
104b
PREDICTION PARAMETER CODING SECTION
CODED PARAMETERS

FIG.5

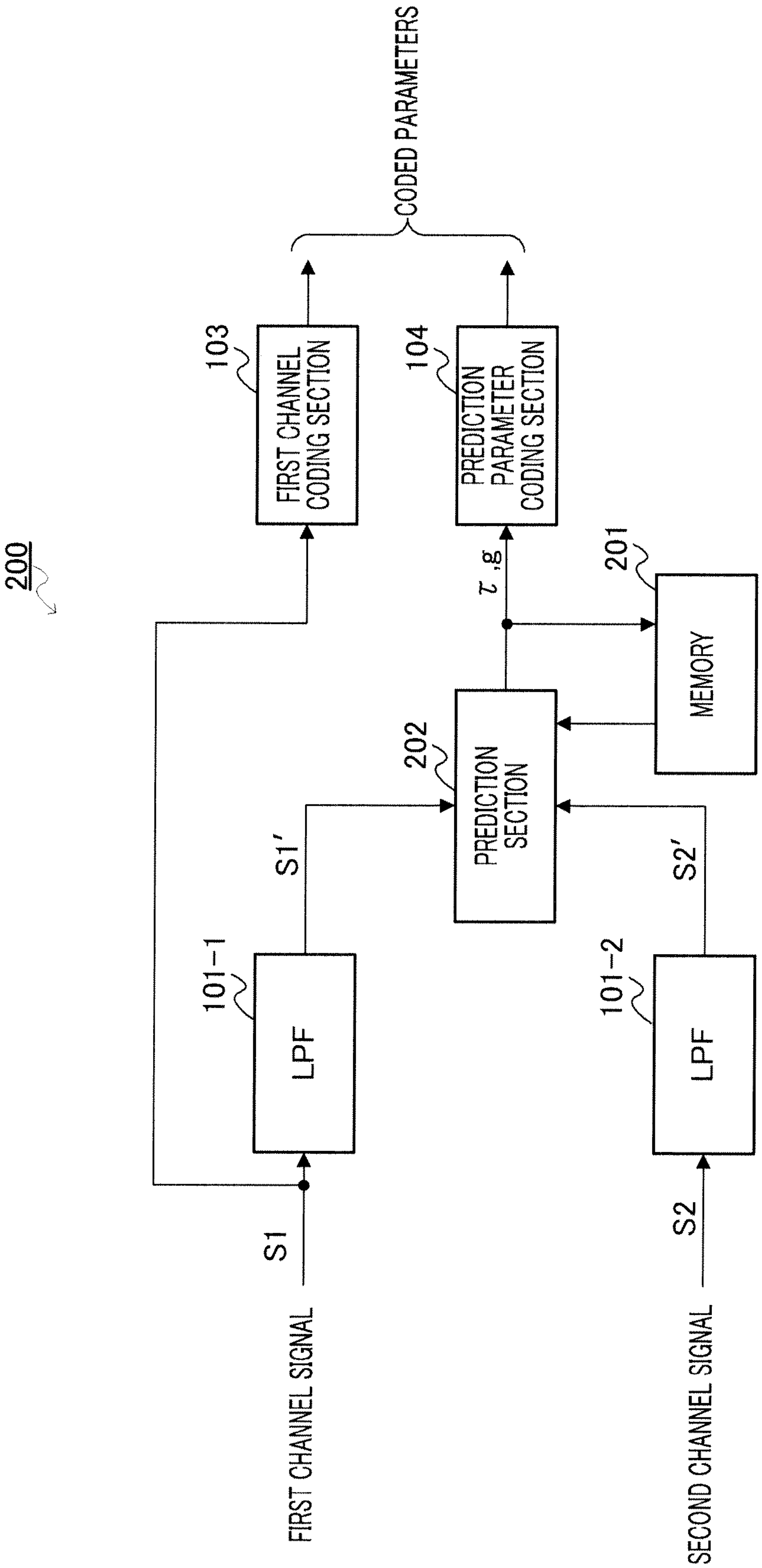
200
CODED PARAMETERS
103
FIRST CHANNEL CODING SECTION
104
PREDICTION PARAMETER CODING SECTION
τ ,g
201
MEMORY
202
PREDICTION SECTION
S1′
S2′
101−1
LPF
101−2
LPF
S1
S2
FIRST CHANNEL SIGNAL
SECOND CHANNEL SIGNAL

FIG.6

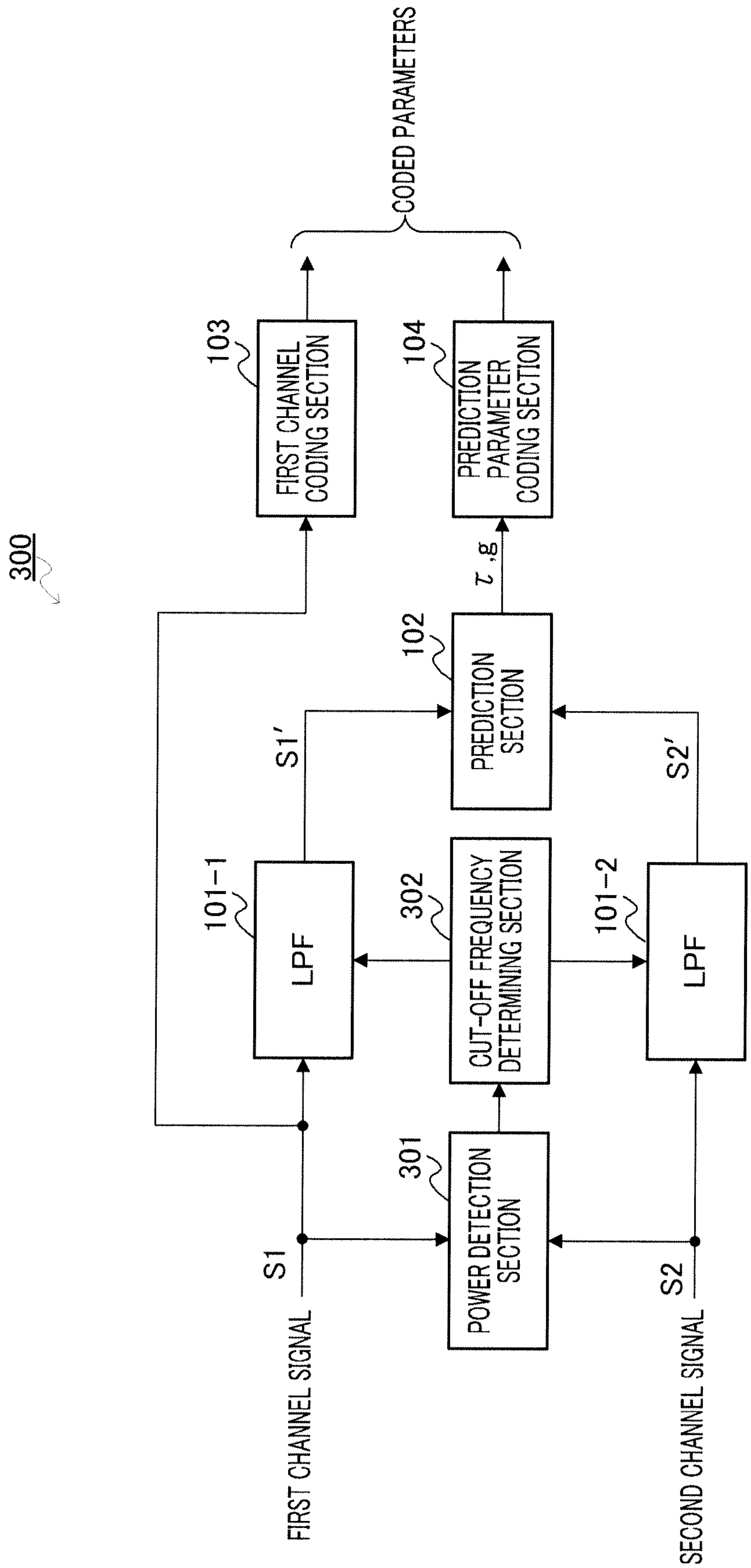
300
FIRST CHANNEL SIGNAL
S1
SECOND CHANNEL SIGNAL
S2
301
POWER DETECTION SECTION
302
CUT-OFF FREQUENCY DETERMINING SECTION
101-1
LPF
101-2
LPF
S1′
S2′
102
PREDICTION SECTION
τ ,g
103
FIRST CHANNEL CODING SECTION
104
PREDICTION PARAMETER CODING SECTION
CODED PARAMETERS

FIG.7

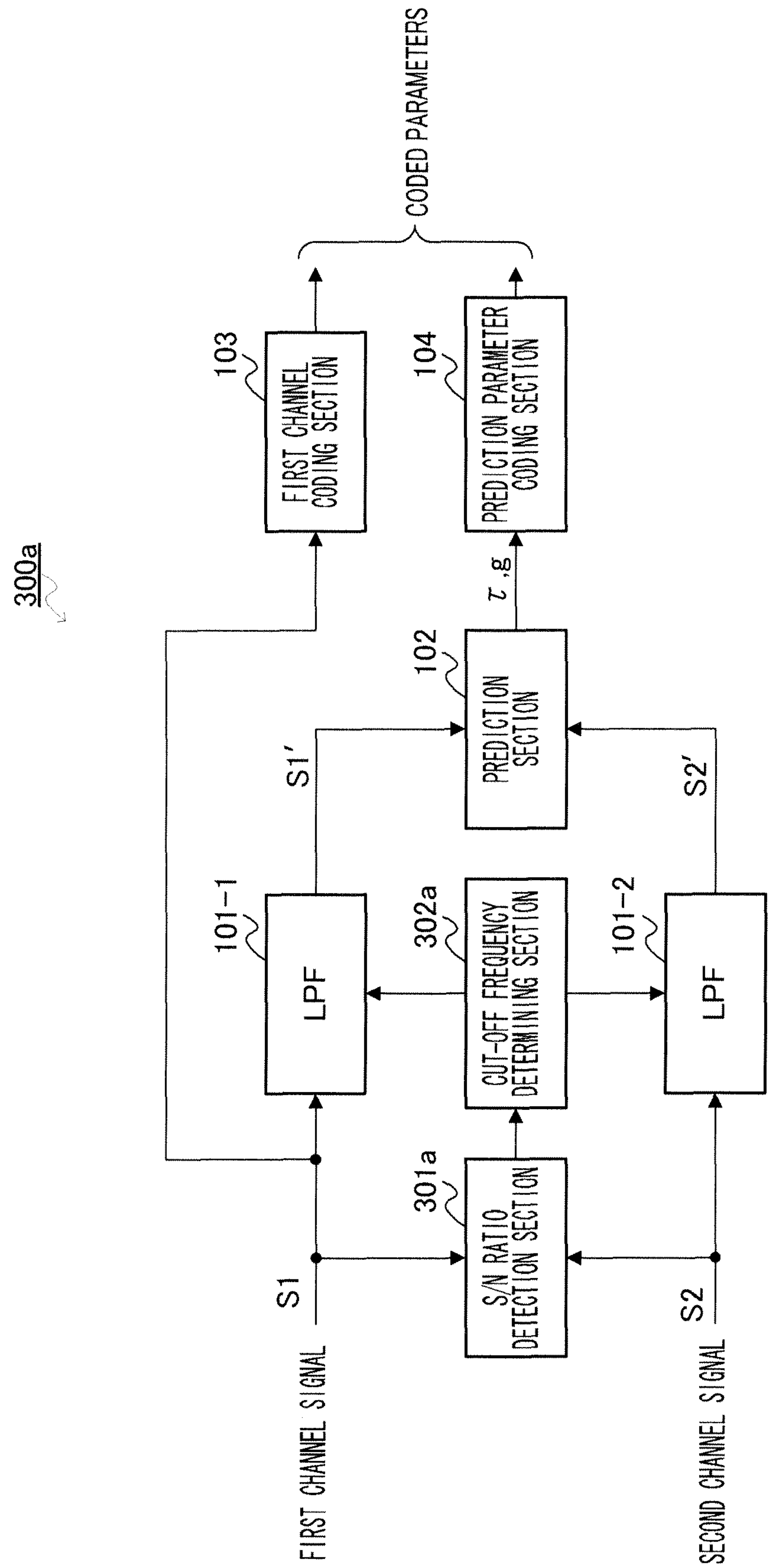
300a
CODED PARAMETERS
103
FIRST CHANNEL CODING SECTION
104
PREDICTION PARAMETER CODING SECTION
τ ,g
102
PREDICTION SECTION
S1′
S2′
101-1
LPF
302a
CUT-OFF FREQUENCY DETERMINING SECTION
101-2
LPF
301a
S/N RATIO DETECTION SECTION
S1
S2
FIRST CHANNEL SIGNAL
SECOND CHANNEL SIGNAL

FIG.8

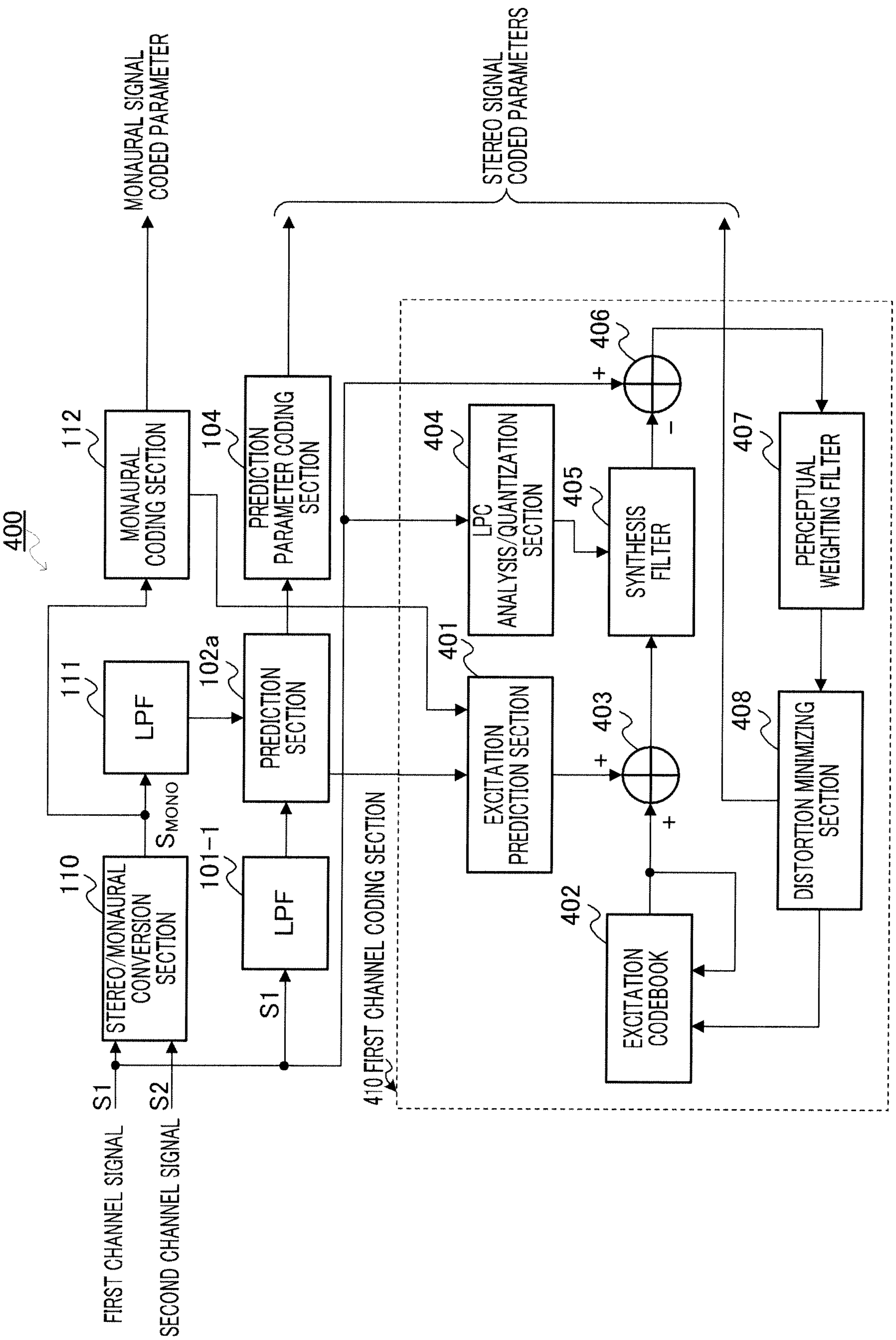
400
FIRST CHANNEL SIGNAL
SECOND CHANNEL SIGNAL
S1
S2
110
STEREO/MONAURAL CONVERSION SECTION
S_MONO
111
LPF
112
MONAURAL CODING SECTION
MONAURAL SIGNAL CODED PARAMETER
101-1
LPF
S1
102a
PREDICTION SECTION
104
PREDICTION PARAMETER CODING SECTION
STEREO SIGNAL CODED PARAMETERS
410 FIRST CHANNEL CODING SECTION
401
EXCITATION PREDICTION SECTION
402
EXCITATION CODEBOOK
403
404
LPC ANALYSIS/QUANTIZATION SECTION
405
SYNTHESIS FILTER
406
407
PERCEPTUAL WEIGHTING FILTER
408
DISTORTION MINIMIZING SECTION
+
+
+
-

FIG.9

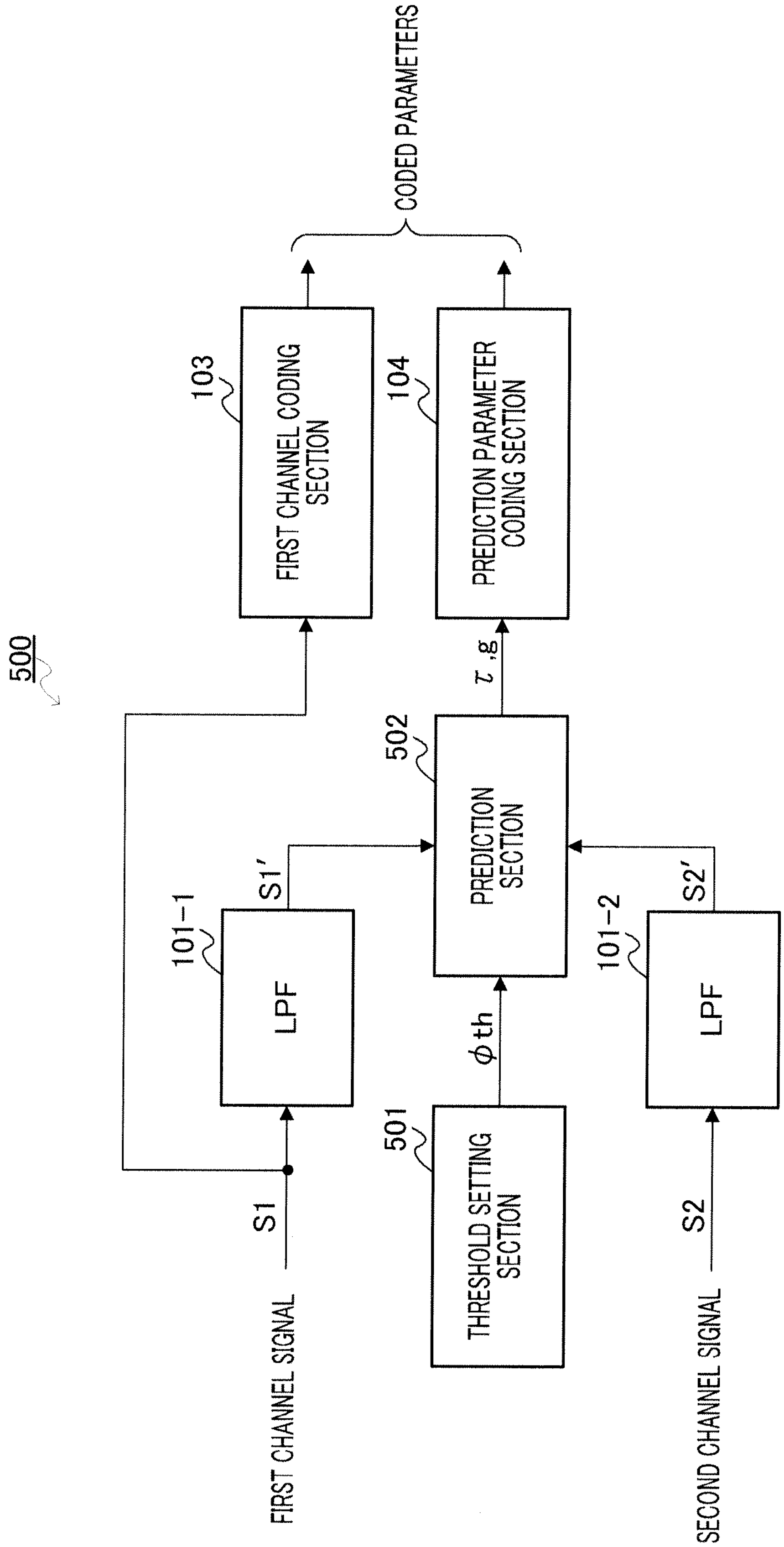
500
FIRST CHANNEL SIGNAL
S1
101-1
LPF
S1′
SECOND CHANNEL SIGNAL
S2
101-2
LPF
S2′
501
THRESHOLD SETTING SECTION
φth
502
PREDICTION SECTION
τ, g
103
FIRST CHANNEL CODING SECTION
104
PREDICTION PARAMETER CODING SECTION
CODED PARAMETERS

FIG.10

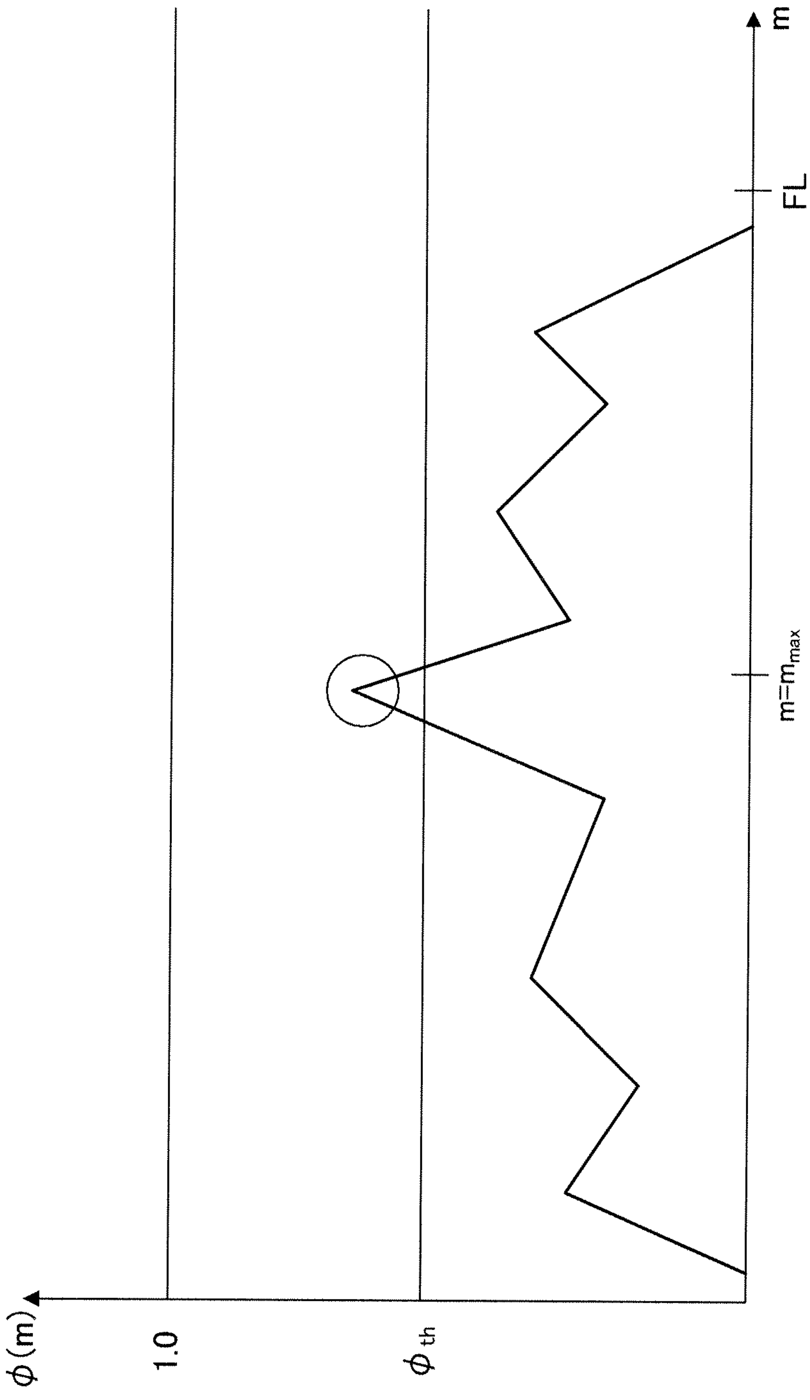

FIG.11

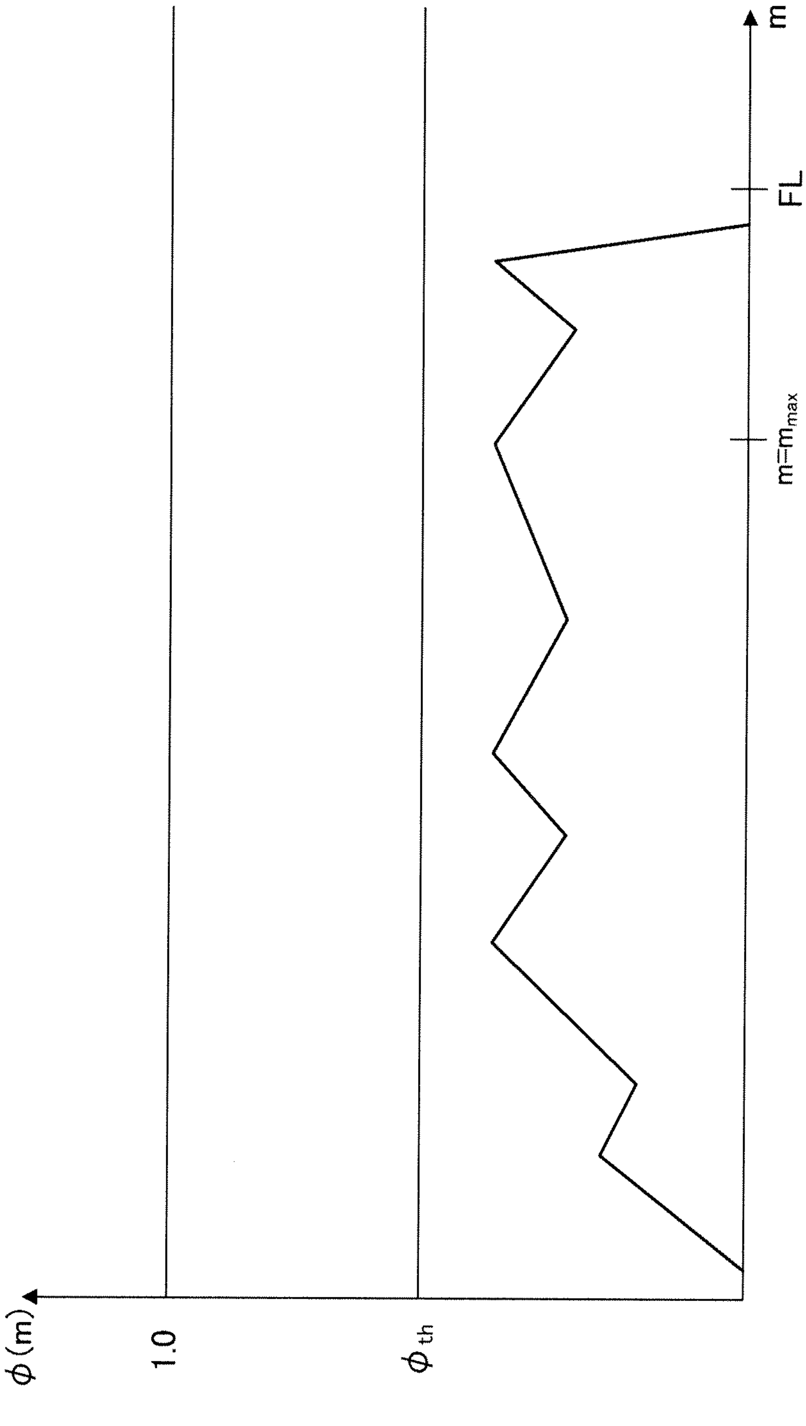

FIG.12

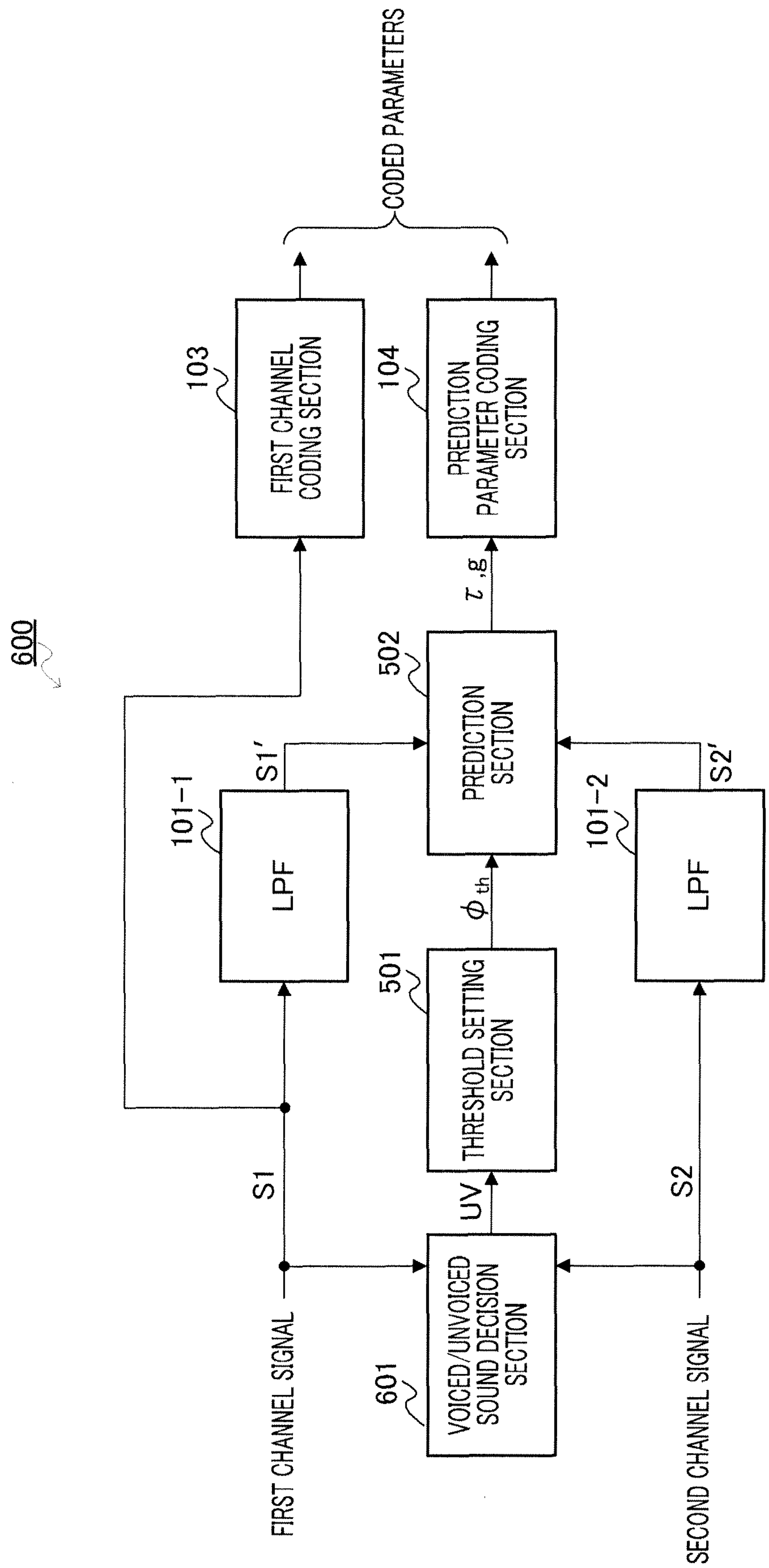
600
CODED PARAMETERS
103
FIRST CHANNEL CODING SECTION
104
PREDICTION PARAMETER CODING SECTION
τ ,g
502
PREDICTION SECTION
S1′
101−1
LPF
S2′
101−2
LPF
φ th
501
THRESHOLD SETTING SECTION
S1
UV
S2
601
VOICED/UNVOICED SOUND DECISION SECTION
FIRST CHANNEL SIGNAL
SECOND CHANNEL SIGNAL

FIG.13

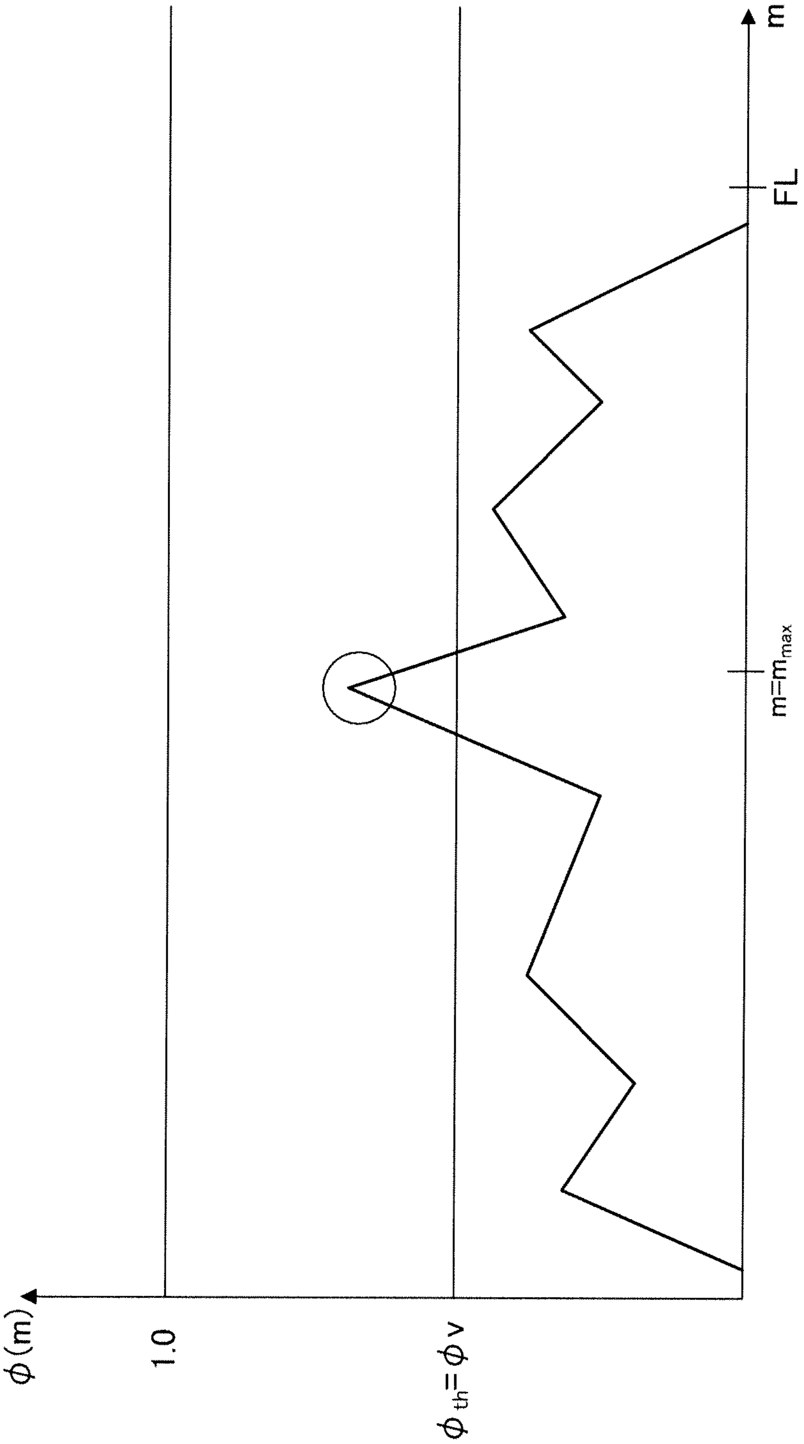

FIG.14

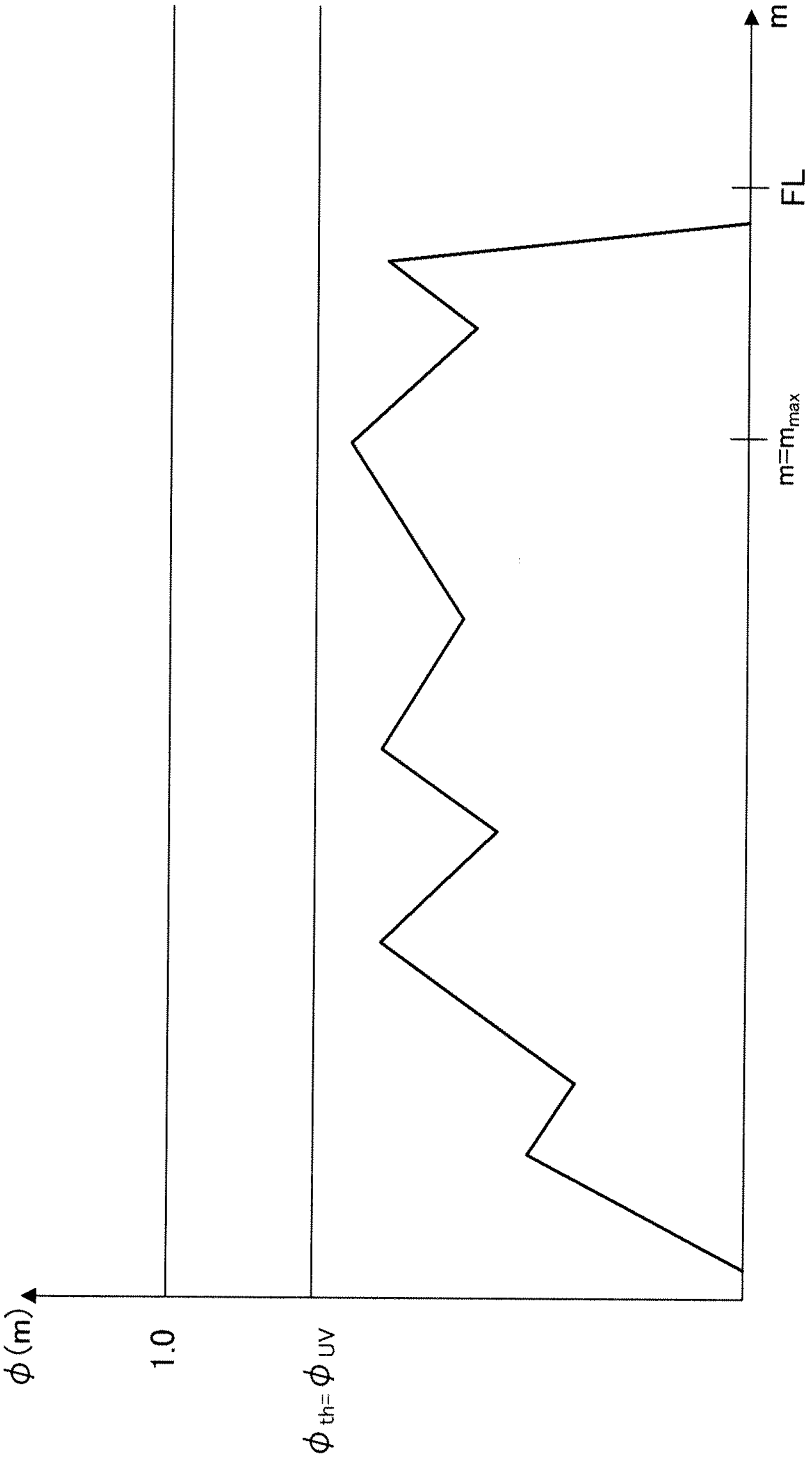

FIG.15

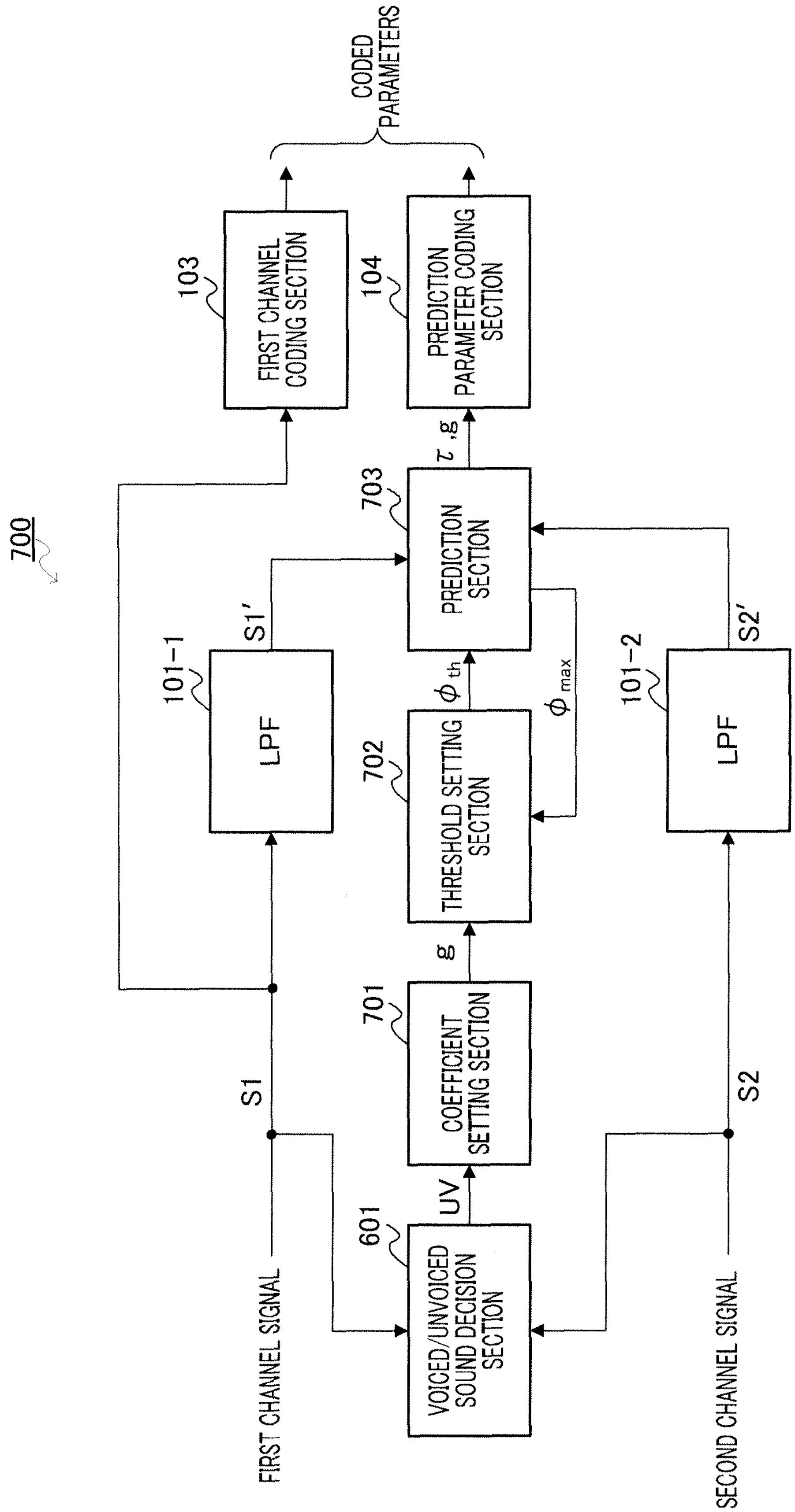
700
FIRST CHANNEL SIGNAL
SECOND CHANNEL SIGNAL
S1
S2
601
VOICED/UNVOICED SOUND DECISION SECTION
UV
701
COEFFICIENT SETTING SECTION
g
702
THRESHOLD SETTING SECTION
ϕ_{th}
ϕ_{max}
703
PREDICTION SECTION
τ ,g
101-1
LPF
S1′
101-2
LPF
S2′
103
FIRST CHANNEL CODING SECTION
104
PREDICTION PARAMETER CODING SECTION
CODED PARAMETERS

FIG.16

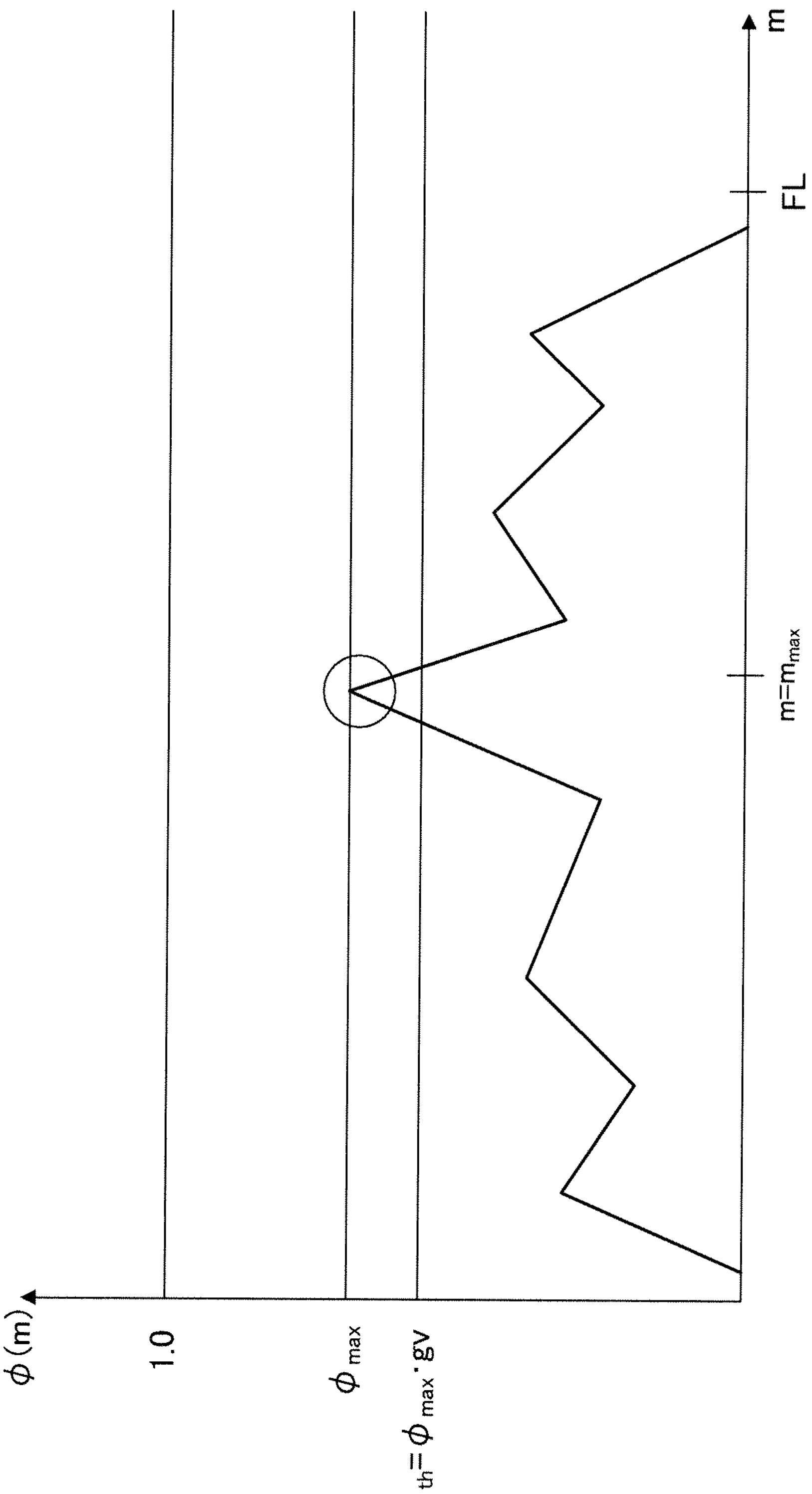

FIG.17

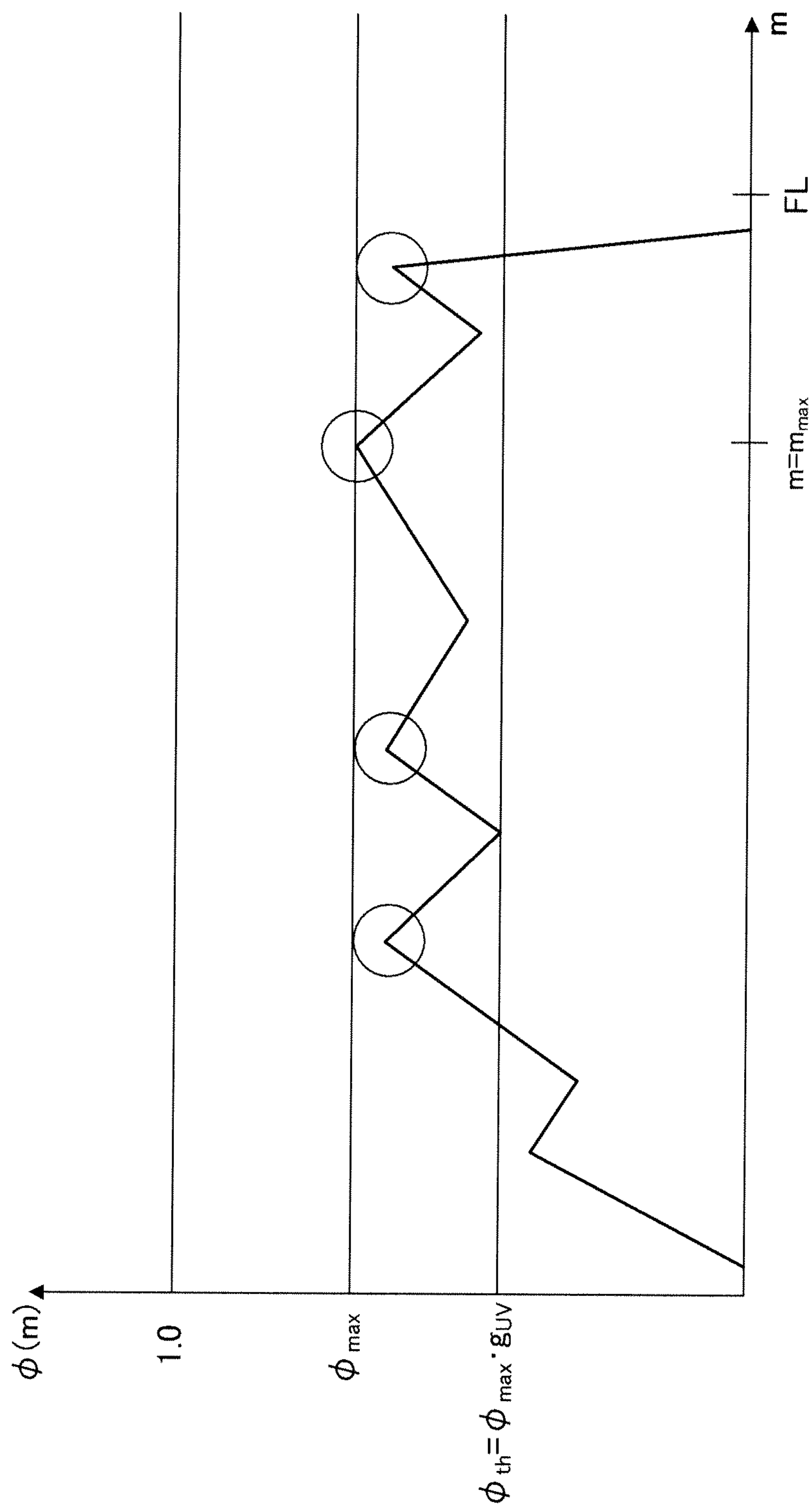

FIG.18

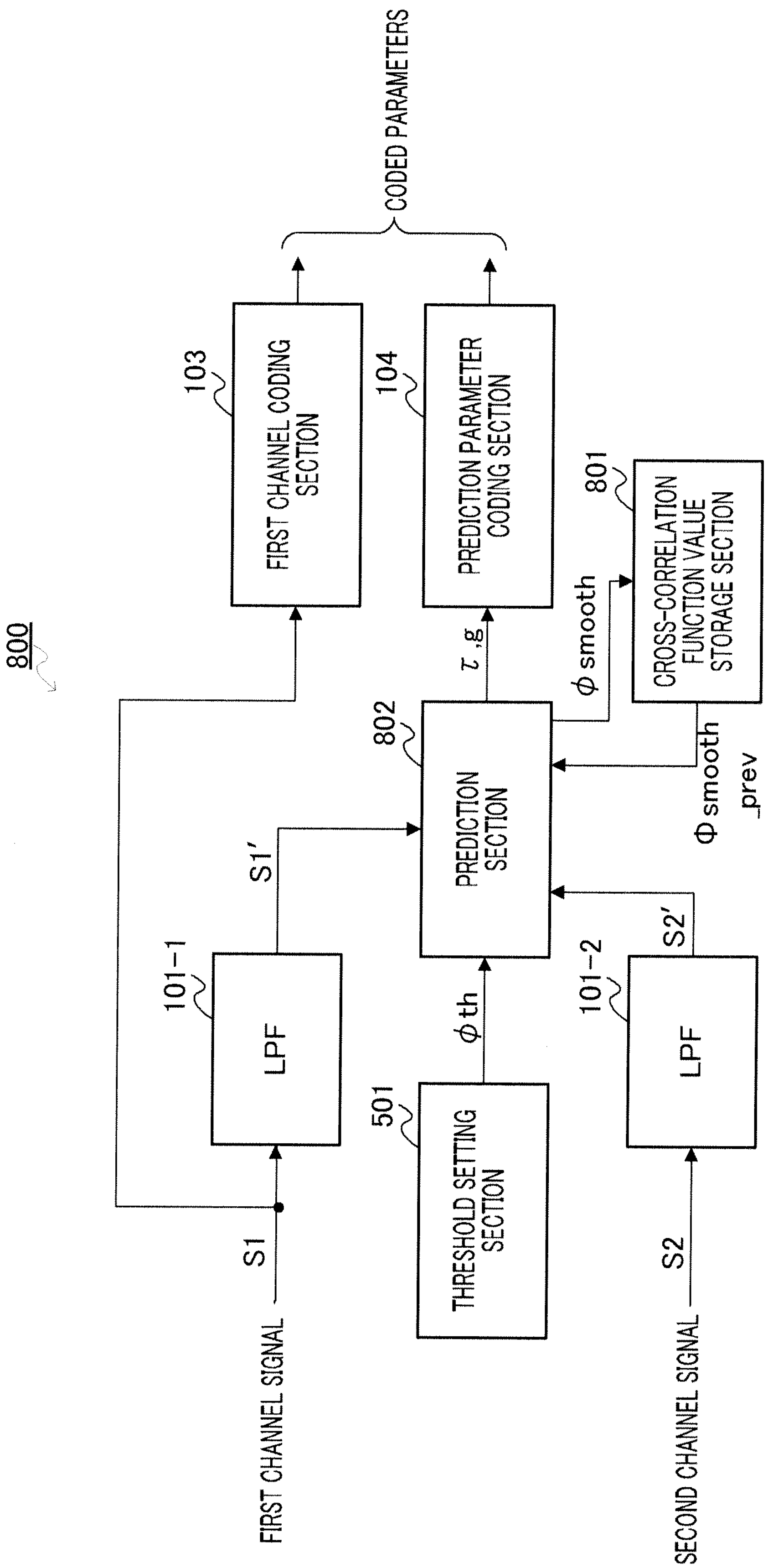
800
FIRST CHANNEL SIGNAL
S1
LPF
101-1
S1′
SECOND CHANNEL SIGNAL
S2
LPF
101-2
S2′
THRESHOLD SETTING SECTION
501
φ th
PREDICTION SECTION
802
τ ,g
φ smooth
Φ smooth _prev
CROSS-CORRELATION FUNCTION VALUE STORAGE SECTION
801
FIRST CHANNEL CODING SECTION
103
PREDICTION PARAMETER CODING SECTION
104
CODED PARAMETERS

FIG.19

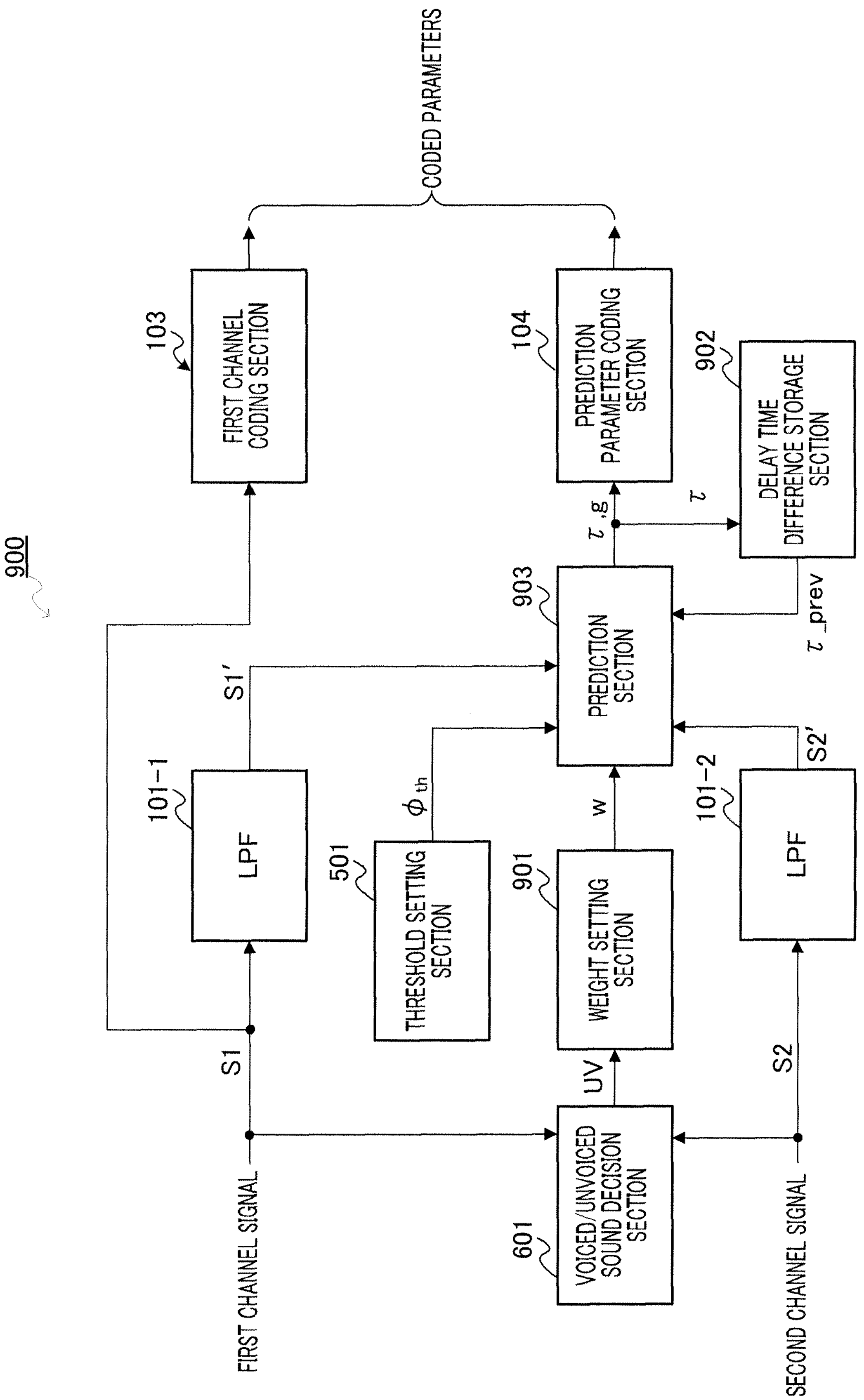
900
FIRST CHANNEL SIGNAL
SECOND CHANNEL SIGNAL
S1
S2
601
VOICED/UNVOICED SOUND DECISION SECTION
UV
501
THRESHOLD SETTING SECTION
901
WEIGHT SETTING SECTION
101-1
LPF
101-2
LPF
φ th
w
S1'
S2'
903
PREDICTION SECTION
τ _prev
τ ,g
τ
902
DELAY TIME DIFFERENCE STORAGE SECTION
103
FIRST CHANNEL CODING SECTION
104
PREDICTION PARAMETER CODING SECTION
CODED PARAMETERS

FIG.20

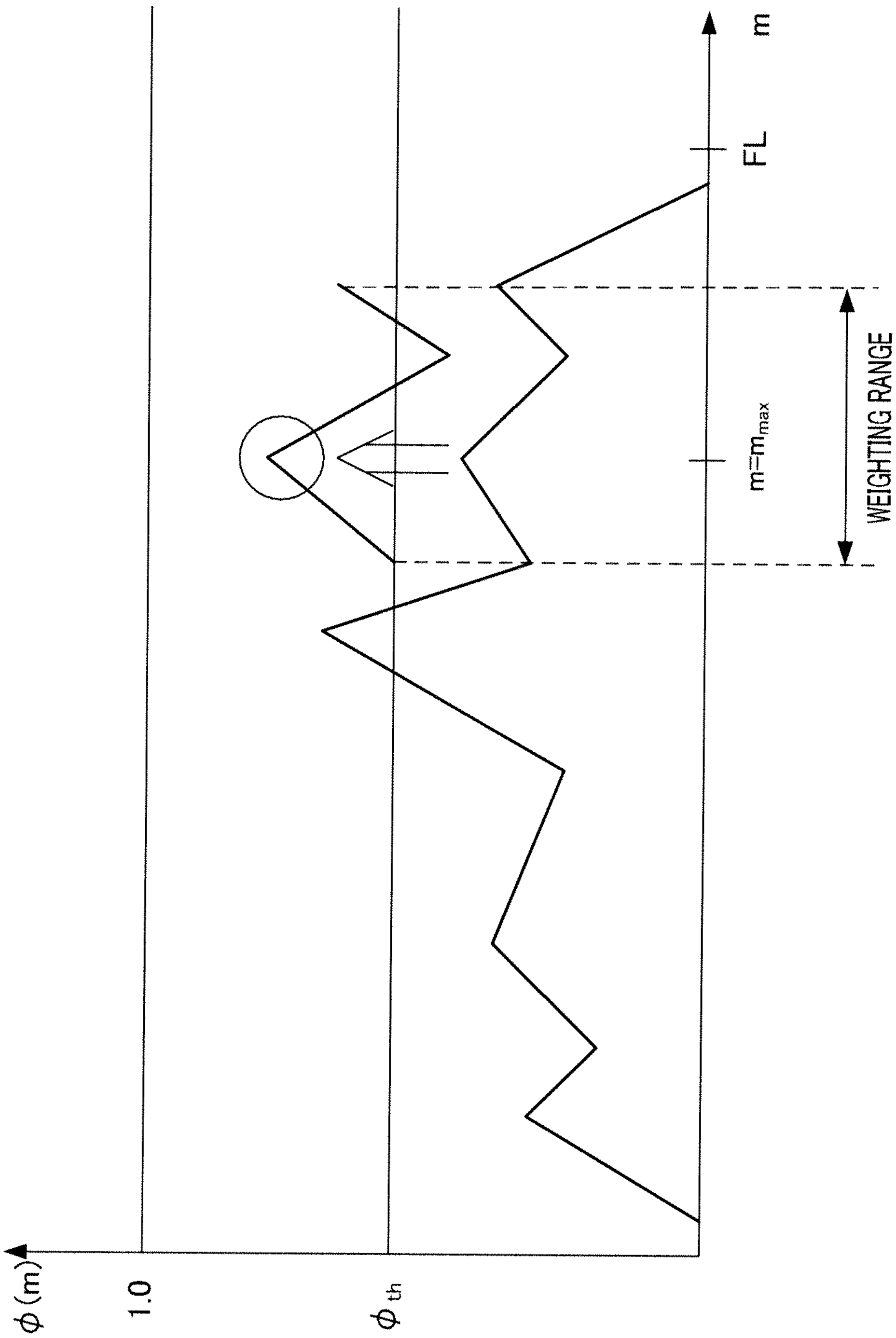

FIG.21

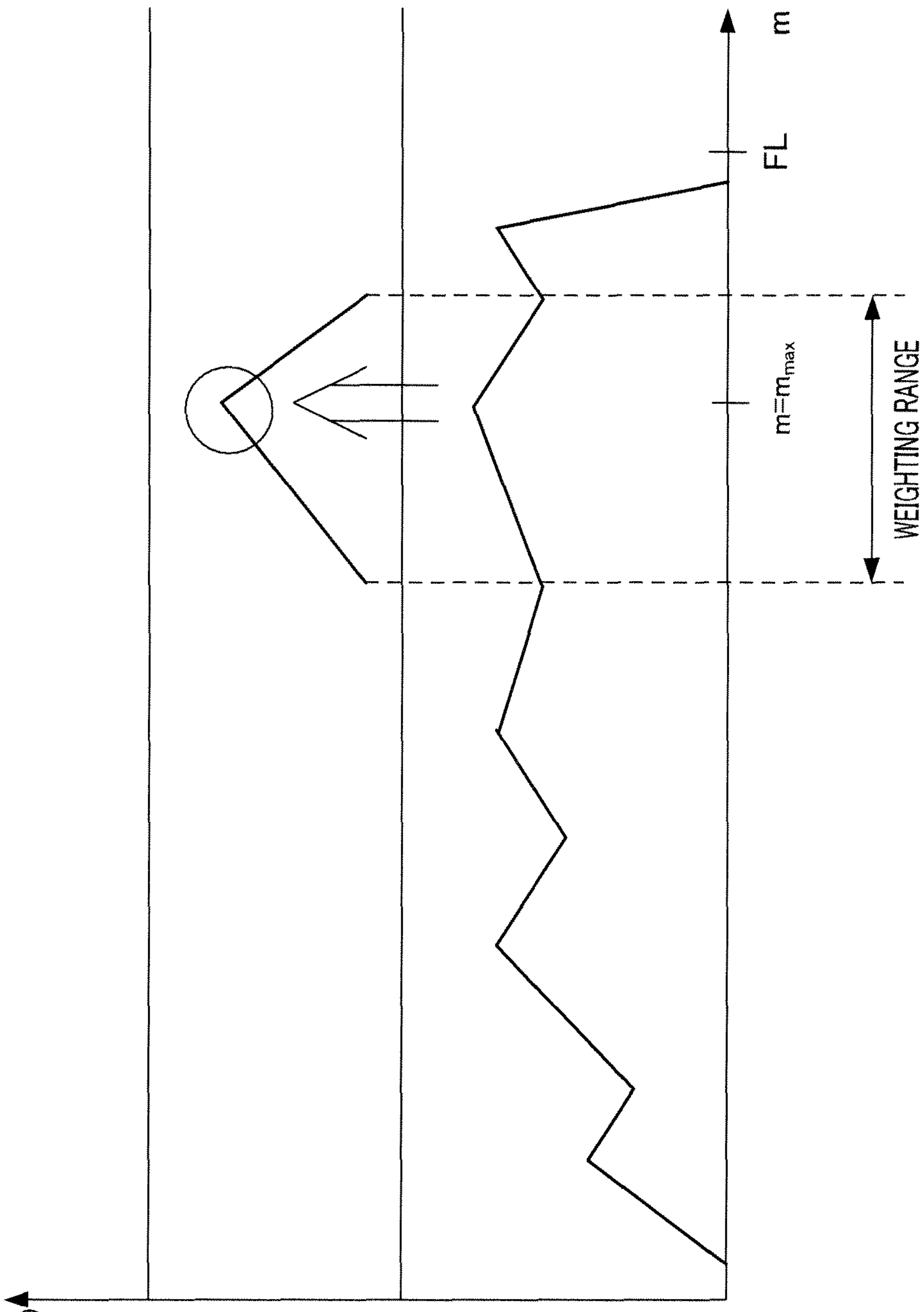

FIG.22

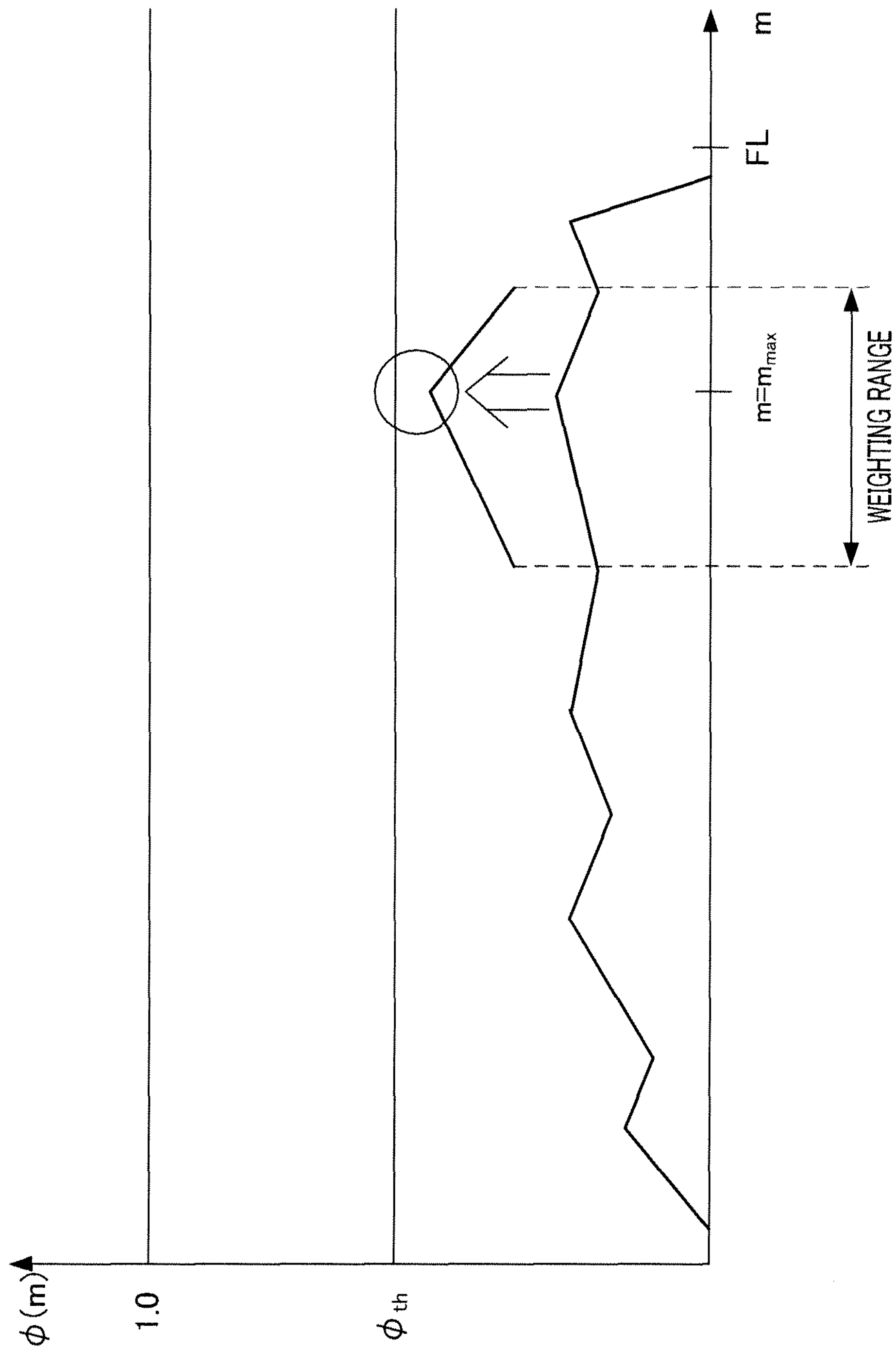

FIG.23

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2006/321673

A. CLASSIFICATION OF SUBJECT MATTER
G10L19/00(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G10L19/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2006
Kokai Jitsuyo Shinan Koho	1971-2006	Toroku Jitsuyo Shinan Koho	1994-2006

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
IEEE Xplore, JSTPlus(JDream2)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y A	JP 7-87033 A1 (Sharp Corp.), 31 March, 1995 (31.03.95), Full text; all drawings (Family: none)	1,2,6-8, 15-17 3-5,9-14
Y A	GOTO et al., "Channel-kan Joho o Mochiita Onsei Tsushinyo Stereo Onsei Fugoka Hoho no Kento", 2005 Nen The Institute of Electronics, Information and Communication Engineers Sogo Taikai Koen Ronbunshu, D-14-2, 07 March, 2005 (07.03.05), page 119	1,2,8,15-17 3-7,9-14
Y A	GOTO et al., "Onsei Tsushinyo Scalable Stereo Onsei Fugoka Hoho no Kento", FIT2005 (Dai 4 Kai Forum on Information Technology) Koen Ronbunshu, G-017, 22 August, 2005 (22.08.05), pages 299 to 300	6,7 1-5,8-17

☐ Further documents are listed in the continuation of Box C.   ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search 27 November, 2006 (27.11.06)	Date of mailing of the international search report 05 December, 2006 (05.12.06)
Name and mailing address of the ISA/ Japanese Patent Office Facsimile No.	Authorized officer Telephone No.

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2006/321673

Claim 3 describes "on the basis of said prediction parameter of the past, a prediction parameter within a predetermined range is generated with reference to said prediction parameter", but fails to specify the relation between a past prediction parameter and a predetermined range clearly, and leaves it indefinite. What is disclosed in the meaning of PCT Article 5 is the description of Par. Nos. [0039] - [0043] and so on, for example.

Claim 4 describes "on the basis of the powers of said first channel signal and said second channel signal, the interruption frequencies of said first low-pass filter and said second low-pass filter are decided", but fails to specify the relation between the value of the power and the interruption frequency decided, and leaves it indefinite. What is disclosed in the meaning of PCT Article 5 is the description of Par. Nos. [0045] - [0050] and so on, for example.

Claim 5 describes "on the basis of the S/N ratios of said first channel signal and said second channel signal, the interruption frequencies of said first low-pass filter and said second low-pass filter are decided", but fails to specify the relations between the values of the S/N ratios and the interruption frequency decided, and leaves them indefinite. What is disclosed in the meaning of PCT Article 5 is the description of Par. No. [0051] and so on, for example.

Claim 10 describes "on the basis of the decision result, said threshold value is set", but fails to specify the relation between the decision result and the set threshold value, and leaves it indefinite. What is disclosed in the meaning of PCT Article 5 is the description of Par. Nos. [0071], [0072] and so on, for example.

Claim 14 describes "according to the number of local peaks contained within a predetermined range from the maximum of said mutually correlated function, said delay time difference is set", but fails to specify the relation between the number of the local peak and the set delay time difference, and leaves it indefinite. What is disclosed in the meaning of PCT Article 5 is the description of Par. Nos. [0076] - [0080] and so on, for example.

Hence, the search or the like has been made on the range supported by and disclosed in the specification, that is, the specific stereo encoding device or the like specifically described in the specification.

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- JP 2005316754 A **[0113]**
- JP 2006166458 A **[0113]**
- JP 2006271040 A **[0113]**

Non-patent literature cited in the description

- **HENDRIK FUCHS.** Improving Joint Stereo Audio Coding by Adaptive Inter-Channel Prediction. *Applications of Signal Processing to Audio and Acoustics, Final Program and Paper Summaries, 1993 IEEE Workshop,* 17 October 1993, 39-42 **[0003]**